EDUCATIONAL AND PSYCHOLOGICAL
MEASUREMENT

Volume II APRIL, 1942 Number 2

A Test for Primary Business Interests Based on a Functional Occupational Classification 113
Alfred J. Cardall

Measurement in Rural Housing
A Preliminary Report 139
Charles I. Mosier

Procedure for Handling Tests and Examinations 153
John V. McQuitty

Machines in Civil Service Testing 167
Sidney W. Koran

Predictive Value of Certain "Law Aptitude" Tests 201
E. L. Welker and T. W. Harrell

An Exploratory Study of Social Guidance at the College Level 209
Margaret Glockler Aldrich

New Tests 217

Measurement Abstracts 220

Measurement News 227

Copyright, 1942, by
SCIENCE RESEARCH ASSOCIATES

PRINTED IN THE UNITED STATES OF AMERICA

A TEST FOR PRIMARY BUSINESS INTERESTS BASED ON A FUNCTIONAL OCCUPATIONAL CLASSIFICATION

ALFRED J. CARDALL
Boston University

AS VOCATIONAL GUIDANCE leaves the area of pleasant gestures and advances towards a realistic process, it becomes more and more dependent upon better methods of evaluating an individual's interests and potentialities. In spite of many imperfections, psychological tests still constitute our best means of diagnosis. It is questionable, however, if even the best analysis of an individual's aptitudes, special abilities, and personality traits helps materially in indicating an occupational area in which the individual will find stimulation, economic independence, and satisfaction in his work. The results of such tests in the hands of a careful counselor serve chiefly in the determination of the "vocational risk" involved in the pursuit of those occupational activities contemplated by the individual.

Perhaps the first step in vocational adjustment should be the crystallization of those interests in the individual as they relate to activities integral with a given job. Investigators in the field of measurement have long been concerned with vocational interests, but no previous attempt has been made to focus specific job-activity preferences on an occupational pattern. It is these data, after all, which point to the initial job, determine the individual's interest or boredom in his first activities, and determine to some extent his progress in it.

Interest measurement has been confined largely to the matter of general interests, and although such interests may be suggestive of occupational areas to be considered, they cannot be regarded as motivators of initial occupational activity.

There is great need today for more specific measurement which will indicate initial jobs compatible with an individual's interest in specific activities.

The construction of the *Primary Business Interests* test is direct, functional, and highly specific in an area of available business positions for beginners. The individual is asked to express his preference or dislike for those specific job-activities which are characteristic of such beginning jobs. Such activities have not been empirically selected, but they were based on an extensive analysis of beginning positions, and only those specific activities which determine and differentiate definite occupational patterns were retained.

Unfortunately no such functional occupational classifications are available — a fact which may explain why so obvious and direct an approach to interest measurement has not been used before. Actually no material changes in the matter of item selection have occurred in 20 years, and the first inventory appearing in 1921 under the name of the *Carnegie Interest Inventory* has set the pattern of which practically all general inventories now available are merely revisions. Scoring methods, it is true, have become highly refined, but it is doubtful if statistical refinement in scoring is any substitute for item-validity or basic data.

Obviously the most desirable way of selecting items for any instrument would be to set the initial research so that the items would evolve as a matter of research rather than empirical choice.

With this in mind, an extensive study was made of initially available business jobs. This study was based on a classroom assignment given to first-year evening college students who had several months' experience in business. Each student kept a work diary of a typical week and was given a grade based on the specificity of detail contained. 106 different jobs covering a considerable range of activity were analyzed, and over 2000 specific items were listed. Items which occurred less than five times in the data were eliminated. Remaining items were reduced to terse expressions of the actual activity. Reduction

in number of items was first achieved by grouping on a basis of a standard terminology wherever possible. Further reduction was based on the concomitance, similarity, and simultaneity of occurrences. The question of whether two items implied the same activity or invariably occurred together was decided by the majority opinion of five vocational counselors. Judgments were expressed independently on forms provided for that purpose. All counselors were actively engaged in placement work.

The list of job-items which appeared in the job-analysis form was the result of the early work in this study. The data used in its construction now have no further significance. The job analysis was used a year later with a different group of evening students who were similarly employed and similarly motivated. As the job analysis was in the form of a check-list, the data appeared in easily tabulated form. A page was attached to cover the listing of any item which may have occurred on the job, for which a printed item was not provided. However, analysis of these additions did not reveal frequencies high enough to warrant their inclusion in subsequent statistical work.

A word of explanation as to the check columns on this form may be in order, although the instructions clearly cover them. These instructions were the same as those used for the final form and are given on page 132. The first set of columns provided a quantitative analysis of items occurring on the job. The instructions given for the second set of columns, however, were designed primarily to motivate the individual. Actually these columns had a definite research purpose to fulfill. The information they provided as to how often the items occur in the job ahead was an important factor in determining their ultimate importance. Nearly 300 job analyses were received, but only 245 of these covered positions initially available in the field of business. But before proceeding with the statistical analysis of these data, let us first review what has been done within the field of occupational classifications.

We have indicated the soundness of an interest test based

on an individual's preference for specific job-activities. In order to score such a test, however, we are concerned with the way and manner in which these activities may be grouped into occupational patterns.

Occupational Classifications

The various methods of occupational classifications are largely empirical in nature. The majority are based on the census classification. This grouping is primarily concerned with economic factors. The importance of such a classification cannot be overlooked since a large part of the statistics available are based upon it.

From the point of view of similarity of activities, the census classification is of little use. To associate musicians and osteopaths, or showmen and college presidents under the same heading gives no clue as to functions involved. Under trade, to follow advertising agencies by stock-yards is similarly of little help in understanding the nature of the activities. A further weakness is illustrated by listing together inventors and draftsmen under professional service; no concern is evidenced for occupational levels.

With these difficulties in mind, an improvement on this classification was reported by Edwards[1] at a recent meeting of the American Statistical Association.

A more elaborate outline was presented at the same meeting by Palmer.[2]

Kimball[3] redistributes the number of gainfully employed as reported in the census figures, by percentages. Reduced as it is to socio-economic groups, his study is serviceable in showing certain shifts of employment.

Similar classifications arise to expedite the very important

---

[1] Alba M. Edwards, "A Social-Economic Grouping of the Gainful Workers of the United States," *Journal of the American Statistical Association,* XXVIII (Oct. 26,1940), p. 378.

[2] Gladys L. Palmer, "The Convertibility List of Occupations and the Problems of Developing It," *Journal of the American Statistical Association,* XXXIV (1939), p. 700.

[3] Bradford F. Kimball, *Changes in the Occupational Pattern of New York State.* (Albany, New York State Education Department. Educational Research Studies No. 2, 1937), p. 38.

matter of recording and tabulating job placements. For example, the Massachusetts State Employment Service classifies their placements by industrial groups, very similar to the census classifications, as well as by occupational groups.

The *Dictionary of Occupational Titles* divides the major occupational groups into seven classifications, arranged alphabetically and identified by the first and second digits of the code numbers. Job classifications within these major groups are identified by three digit groups following the first two code numbers.

Humphreys[4] gives us another classification which is concerned with the general functional aspects of jobs as they apply to many industrial and commercial establishments. This classification groups functional activities regardless of the industry or field in which they are found.

The counselor is more concerned with a similarity of the clerical functions in different fields than with the differences between the fields themselves. It is the actual function of the job which is significant to the prospective worker. This concern with the worker leads to other methods of classifications. Kelley[5], Thurstone[6] and others would classify occupations by the pattern of abilities required. This concern with multiple abilities will have far-reaching effects in occupational choices on the basis of profile-matching if the vocational application of such profiles is ever understood.

Kitson[7] attempts to group vocations by the kind of training required, and it is to be regretted that this approach has not been followed up by more specific work based on classifications of pre-entry requirements quantitatively expressed.

Brewer[8] gives us a three-dimensional concept of occupations classified by fields, functions, and occupational levels.

[4] J. A. Humphreys, *How to Choose a Career.* (Chicago, Science Research Associates, 1940), 48 pp.

[5] Truman L. Kelley, *Essential Traits of Mental Life.* (Cambridge, Harvard University Press, 1935).

[6] L. L. Thurstone, "A Multiple Factor Study of Vocational Interests," *Personnel Journal,* No. 10 (Oct. 1931), 198-205.

[7] Harry Dexter Kitson, *The Psychology of Vocational Adjustment.* (Philadelphia, J. B. Lippincott Co., 1925).

[8] John M. Brewer, *Occupations.* (Boston, Ginn & Company, 1936), 437-441; 590-597.

This limited treatment of the various approaches serves to illustrate a variety of classifications including industries, socio-economic factors, intelligence, abilities, and general interests. From the guidance point of view we are not concerned with the socio-economic status, as the number of trained workers in each occupation tells us little of its function. Classifications on the basis of abilities and intelligence give us more definitely an idea of the requirements of occupations, but disregard underlying interest in such activities. Classifications based on broad interest patterns fail too in that such patterns express only a broad attitude rather than immediate and specific interests in the actual work of the beginning occupation.

A method of occupational classification of inestimable value would be one which brought together those jobs which call for the same specific activities in approximately the same proportion, regardless of field or title. In only a very general sense can we assume that either a mention of the field or occupational title gives any indication as to what is inherent in the job itself. In fact, we may regard such descriptions as often adding confusion to an already complicated picture. The only sound basis of grouping these occupations lies in the specific nature of the work, and accordingly calls for the same general structure of interest, skills, and personality traits on the part of the worker. Although the psychometrist has accomplished much from a diagnosis of an individual's personal qualities, he has, as yet, been unable to indicate the social or economic significance of these same qualities. In the last analysis it is the latter phase which makes the first meaningful.

This study presents a statistical approach to such a functional classification but within a limited range of occupational activity. Its purpose is to measure the relationships between specific job-activities of initially available business positions and to discover what common factors exist so that special functions may be isolated. The activities which in themselves are closely related and conversely alienated from the others form the patterns needed as a scoring scheme for our interest test.

The same method of pattern determination is equally ap-

plicable to other threshold positions, as well as various occupational levels. The extension of this work would be infinitely worth while and of far-reaching consequence in the field of guidance.

Setting Up a Contingency Table

Our data consist of 245 analyses of initially available business jobs. The specific activities have been checked in each analysis in such a way as to indicate whether the activity occurs *much* or *occasionally* in the job. Before a computation of the interrelationships of these items can be made, a tabulation of the number of times that each item occurs in the data as well as the number of times which each item occurs with every other item must be recorded. That is to say, we must know how often item No. 1 occurs, also how often it occurs with 2, 3, 4, 5, and up to 115. We must know how many times item No. 2 occurs with 3, 4, 5, etc., and similarly until we have a table of all such contingencies.

A contingency table so constructed gives us at a glance a frequency of concomitance of any item with all others. Since this particular table comprises 6,550 cells above the diagonal, space hardly permits its inclusion here. The diagonal values themselves represent the number of times that each item occurs in the 245 analyses and the largest value that any can take is, of course, 245.

In constructing this table the I.B.M. tabulating equipment was used, a Hollerith card being punched for each case.

The card provides 80 columns which may be punched from zero to nine in each, and above this range of digits two other positions may be punched, making in all twelve positions within a column. In this instance, the first three columns were used to carry the number of the job-analysis form. For example, form 133 has 1 punched in the first column, 3 in the second, and 3 in the third. In the fourth column the items between 1 and 9 occurring as either *much* or *occasionally* on the job-analysis were punched; in the fifth column number 10 was punched as zero, 11 as 1, up to 19 as 9; in the sixth column

items 20 to 29; and similarly for the remaining. The top two positions were not used since dichotomies of ten items in each column facilitated rapid punching and eliminated more complex conversions. After cards were punched, they were checked by another clerk. The resultant contingencies formed the basis for succeeding statistical work.

Weighting the Job-Items

Our contingency table, comprising some 6500 cells, gives us graphically the raw count of the occurrence of each item with every other. We cannot assume, however, that all items are of equal importance. Before going further with the not inconsiderable computational work of pattern-determination, let us see what items might now be eliminated as of little importance in resultant groupings. What factors are significant in making this decision?

Obviously, the frequency of occurrence of the item, indicated as a diagonal value, must be considered; so, too, its variance is a natural weighting factor. Other *á priori* considerations also occur which may or may not be inherent in the data. Of a list submitted to a group of placement officers these four were considered most important:

1. Proportionate time devoted to each item on the job.
2. Need in job-activity of position ahead.
3. Amount of training required previous to employment.
4. Relative importance in the selection of the employee.

The first two of these considerations could be determined from the data, since the job-analysis form used and previously described provided check columns, so that an *M, O* or *R* (much, occasionally, or rarely) response could be recorded in respect to each of these considerations. The third and fourth considerations, however, could not be objectively determined. Five vocational experts were, therefore, asked to rate each job item on a three-point scale in these respects. The following paragraphs are quoted from the written instructions to the

judges, which were further clarified in a group meeting before the ratings were made.

"Amount of training involved. The phrase refers to the amount of training received before employment and considered necessary by the employer to perform the job activities involved. Such training naturally differs in amount, which may be illustrated by such items as "post bookkeeping entries," in which case the employer would expect the individual to have had some training, and such an item as "make up balance sheet" in which case considerable training previous to employment would have been necessary. Please consider carefully, therefore, each of the 115 items in respect to this consideration, using the first column for your check mark if you consider a relatively large amount of training is involved previous to employment and using the second column if you consider a lesser amount involved. Make no check marks if the amount of training previously necessary is relatively little.

"Relative importance in selection of employee. The ability to do certain of these items may be an important consideration in the selection of an initial employee. This is particularly true in respect to job-activities of the contact type. Make no check marks if the ability to do the activity is of relatively little importance in selection. Use the fifth column where ability to do this item becomes of importance in selection and step up your check mark to the fourth position if you feel that it is of considerable importance in employee selection. To illustrate, the question of being able to "call on clients" is undoubtedly a factor in selection. But an item such as "sell goods on commission basis" is considerably more important, the ability to do which is probably the primary consideration in the selection of an employee where such an item probably constitutes his chief activity."

The weighting formula for each item is composed of these six considerations. Since each cell is determined by the concomitant occurrence of two items, it consequently has a weight equal to the product of the weights of the two diagonals which compose it, and no such result, of course, can exceed the product of the square roots of the weighted diagonals; can in fact equal it only when there is maximal concomitance of any two items in our data. The formula for the Aggregate Weight[9] of each cell value is, therefore, composed of the square roots of each frequency, the square root of each variance which in the case of a proportion is pq, and an additive weight of factors just considered for each of the contingent items.

[9] The vital assistance of Professor T. L. Kelley of Harvard in developing the statistical procedures of this study is gratefully acknowledged.

$$(5)\ \text{Agg. Weight}_{12} = (W_1 \sqrt{f_1 p_1 q_1})\ (W_2 \sqrt{f_2 p_2 q_2}) \qquad (1)$$

in which $W_1 = w_a + w_b + w_c + w_d$ and

$$w_a = \frac{M}{M + O}, \qquad (2)$$

$$w_b = \frac{2(pu + su) + (=)}{2(f + pu)}, \qquad (3)$$

$$w_c = \frac{2\,Lt + St}{2j}, \qquad (4)$$

$$w_d = \frac{2\,Cs + Ss}{2j} \qquad (5)$$

Formulae 2 through 5 take on values between 1 and 0 and are additive in their function. These same numbers will identify each formula with the consideration afore-described. The symbols are thus interpreted. In (1) *M* is the number of times the item occurs *much* on the job, and *O occasionally*.

In (2) *pu* refers to the *pick-ups,* or frequencies of occurrence in job ahead but not in present job, while *su* indicates *step-ups* in the amount of the activity in the next job over the present, and = refers to the tabulations of equal amount of activity in job ahead as in present job.

In (3) *Lt* indicates the number of judgments that a large amount of pre-entry training is required, *St* a somewhat smaller amount; *j* the number of judges.

In (4) *Cs* indicates the judgments that ability to perform the job activity is a considerable factor in selection of employee, and *Ss* somewhat of a factor in selection.

These aggregate weights range as low as .44 on item 105 to as high as 8.07. Items 12, 15, 16, 74, 78, 79, 105, 106, 108, and 112 were eliminated because of low weights as definitely of little importance in further consideration. These ten items had weights below unity and/or frequency of 12.

Computing Correlation Coefficients

In order to determine the relationships that existed between these contingent frequencies, the cell values given in

the contingency table were converted into the more usual correlation mold. It may be observed that the raw tabulations in the contingency table can be expressed as p's, proportions, computed by dividing the observed values by 245, which is the total number of times it could have occurred in our sample. This table of p's can next be converted into co-variances by subtracting from the cell p the product of the diagonal p's occurring in its row and column. This p_{12} minus p_1p_2 becomes the numerator of the now easily recognizable formula for the product-moment correlation, the denominator being the standard deviation of each variable, which in this special case is $\sqrt{pq}$ of each. These values are readily obtained from the new Kelley tables[10] for any given p. For the contingency method, then, of computing correlation coefficient we have

$$r_{12} = \frac{p_{12} - p_1p_2}{\sqrt{p_1q_1}\ \sqrt{p_2q_2}}, \qquad (6)$$

in which $p_{12} = \dfrac{f_{12}}{N}$ (cell)

$p_1 = \dfrac{f_1}{N}$ (diagonal)

$q_1 = 1 - p_1$

$N = 245$

The resulting intercorrelation table gives all positive correlation values of .3 or better, and negative correlations of .2 or greater. The dots occurring in other cells indicate smaller values and fall between —.20 and +.30. All values below this diagonal, as in the contingency table, are omitted in the interest of economy, merely duplicating as they do the values above the diagonal.

It will be observed that in only one instance does any value equal .80. This fact would seem to indicate that the work of combining job-items on the basis of concomitance and simultaneity was adequately done. It may be said that a correlation

[10] *The Kelley Statistical Tables,* (New York, The Macmillan Company, 1938).

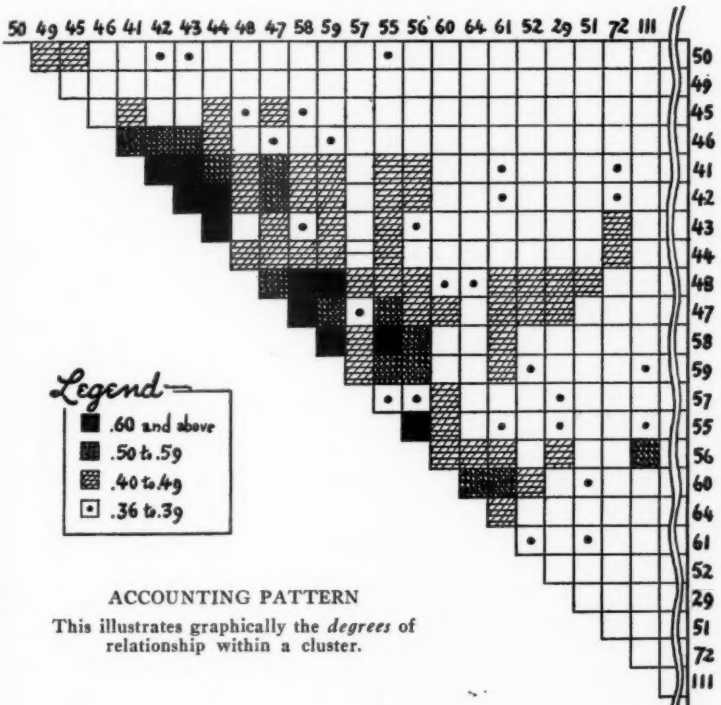
closely approximating unity would be of little value in pattern determination, since it would indicate invariably that one job-activity depended entirely on the other or always occurred with it.

Grouping Correlation Coefficients

Our task now becomes one of discovering what functional patterns of occupational activities exist in our data. The problem of a cluster analysis may be visualized as a plateau of relationships which, if rearranged in rows and columns, would result in a topographical map; high relationships forming a peak surrounded by related and interrelated activities. In a sand-pile graph, low relationships form the valleys and do not help in distinguishing one cluster from another. In short, these correlation coefficients must be so inter-changed within the matrix that the highest value in each row and column appears as near the diagonal as possible. If such patterns as we have postulated are inherent in the data, this re-arrangement should result in clusters along the diagonal of the new matrix.

A table representing this re-arrangement was constructed showing that several clear-cut clusters or patterns are now evidenced. Values of .40 and above occurring in a row and column appear to indicate the extent to which an item "belongs" to a pattern, and very few such values occur very far removed from the diagonal. A closer grouping of these items, however, is possible by the further elimination of such items where no value of .40 or better occurs in its row and column, and where two items, such as 107 and 109, while closely related, appear to be nearly discrete in themselves. Twenty-nine items may now be eliminated as having no value as great as .40, leaving only 75 items which fall into specific groupings. In a very few cases items have values too high to be discarded, but appear to belong equally well to two patterns. In delineating these patterns, therefore, these items will be given one-half weight in respect to each pattern.

This table is too extensive to reproduce here but a single pattern will give a graphic illustration. By converting figures into toned equivalents, the size of a correlation coefficient becomes a density value which illustrates the degree of relationship within patterns.



ACCOUNTING PATTERN

This illustrates graphically the *degrees* of relationship within a cluster.

Following this chart are tables in which each cluster has been treated as a single matrix, so that the relationship between the items within it may be more carefully studied. The name of each pattern has been determined by the most apparent function within it.

TABLE 1

ACCOUNTING

49	45	46	41	42	43	44	48	47	58	59	57	55	56	60	64	61	52	29	51	72	111
	42																				
			42			40		40													
			50	55	51	43															
				80	64	56	45	53	44	46		44	42								
					71	65	45	53	46	48		46	40								
						69		46		47		41								46	
							48	47	44	48		40								42	
								56	63	65	44	46	48			45	42	46	42		
									61	59		52	49	44			47	40			
										79	48	63	59		41						
											47	58	57			41					
														46							
													66	45							
														46	45	41		44			50
															50	50	48				
																43					

50 Verify bookkeeping records, audit, etc.
49 Set up new system of accounts
45 Make up tax returns
46 Keep inventory records
41 Make bookkeeping entries
42 Post bookkeeping entries
43 Take off trial balances
44 Make up balance sheet profit and loss statement
48 Make out and figure payrolls
47 Make out monthly statements, bills, figure extensions
58 Reconcile check book with bank statement
59 Enter checks received in check book
57 Take care of petty cash
55 Draw checks
56 Make out deposits
60 Figure trade discounts, commissions, etc.
64 Figure salesmen's commissions
61 Figure interest
52 Check invoices, prices, discounts and allowances, etc.
29 Determine credit risks
51 Prepare special reports, sales analyses, etc.
72 Type financial statements
111 Take deposits to banks, cash checks, collect bank statements and have checks certified

TABLES 2 AND 3

COLLECTIONS AND ADJUSTMENTS

26	25	27	28	68	
	50				14
	43				26
		36			25
			56		27
				40	28
					68

14 Call on clients
26 Make personal calls for credit information
25 Make personal collection calls
27 Make customer adjustments, smooth out difficulties
28 Handle complaints, investigate
68 Telephone delinquent customers

JUNIOR CLERICAL

69	111	98	99	103	102	101	104	84	100	
52						46				86
			41							69
			51	42						111
			54				40			98
				65	45		44			99
					58	45	51		45	103
						59			42	102
										101
								46	40	104
										84
										100

86 Mail out statements and correspondence
69 Address envelopes, bills, etc.
111 Take deposits to banks, cash checks, etc.
98 Pay bills and bring back receipts
99 Get mail and stamps at post office
103 Take mail, registered mail, and parcels to post office
102 Seal, weigh, and stamp mail
101 Fold letters, circulars for mailing
104 Run errands for employer
84 Assist others, general handyman
100 Take messages, papers, and distribute mail

TABLES 4 AND 5

STENOGRAPHIC — FILING

86	69	71	67	65	66	96	87	72	
	52								86
		51	40						69
				60		45		49	71
									67
					43				65
									66
									96
							46		87
									72

86 Mail out statements and correspondence
69 Address envelopes, bills, etc.
71 Type letters, orders, forms, etc.
67 Answer telephone
65 Take dictation and transcribe
66 Code and type telegrams
96 File orders, letters, bills, reports, trade information
87 Look up information in files, library, etc.
72 Type financial statements

SALES — OFFICE

22	24	62	90	19	91	13	93	97	94	92	
	41										22
		40									24
			42								62
				37							90
					40						19
						56					91
							84	77	53	47	13
											93
											97
											94
											92

22 Make out price sheets
24 Check on competitors' prices, compare quotations
62 Figure quotations
90 Dictate letters, reports, etc.
19 Attend conference with supervisor
91 Organize work, train and supervise others
13 Make up sales contracts
93 Classify orders to size, patterns, salesmen, etc.
97 Keep trade information, credit information, up to date
94 Make forms and charts
92 Purchase merchandise, supplies, equipment

TABLE 6

SALES — STORE

39	34	7	33	37	4	114	5	40	11	31	1	2	20	9	8	6	113	
47																		38
40																		10
	46				46													39
		51																34
		40																7
			45															33
					61	50	47	41	54									37
						45	51	42	53									4
								40										114
								40										5
									42									40
										44								11
											41							31
												46		44	43			1
														41				2
														55	44			20
															65	47	50	9
																49		8
																	42	6
																		113

38 Make up and schedule shipments, etc.
10 Deliver orders to customers
39 Put up mail and telephone orders, fill orders to be shipped
34 Check on quality of goods, examine for defects
7 Take inventories
33 Check and receive incoming supplies, record, etc.
37 Unpack goods, put away and keep storeroom in order
4 Wrap up bundles and packages
114 Sweep floors, empty waste baskets, clean up, etc.
5 Put tags and labels on merchandise
40 Dust shelves, put in order
11 Restock shelves and cases
31 Give information, quote rates over telephone
1 Wait on customers, sell over the counter
2 Sell goods over telephone
20 Letter signs for stock display
9 Set up displays, window trim, etc.
8 Dismantle window displays
6 Arrange display of food stuffs
113 Clean refrigerator, show cases, equipment

Final Occupational Patterns

The foregoing tables comprise the resultant six occupational patterns, namely, *accounting, collections and adjustments, junior clerical, sales-office, sales-store,* and *stenographic-filing.* These correlation coefficients, however, indicate only the "belongingness" of items within their pattern. The relative significance of each item within the pattern is determined by returning again to the table of weights. We now relist these items under their proper pattern headings with their respective weights as determined earlier in this study.

TABLE 7

ACCOUNTING	Weight
3. Verify bookkeeping records, audit, etc.	3.11
5. Determine credit risks	2.97
12. Make up tax returns	2.59
16. Make out monthly statements, bills, figure extensions	4.21
21. Make up balance sheet, profit and loss statement	3.08
23. Make out and figure payrolls	2.60
26. Set up new system of accounts	1.25
27. Figure interest	2.23
29. Make out deposits	2.42
31. Check invoices, prices, discounts, allowances, etc.	4.05
33. Post bookkeeping entries	3.83
35. Prepare special reports, sales analyses, etc.	4.28
37. Enter checks received in check book	1.31
52. Reconcile check book with bank statements	2.37
55. Keep inventory records	4.46
57. Take off trial balances	4.86
63. Make bookkeeping entries	4.63
68. Figure trade discounts, commissions, etc.	3.14
69. Draw checks	2.61
70. Take care of petty cash	3.57
74. Figure salesmen's commissions	1.82
Score ½ on:	
38. Type financial statements	2.14
6. Take deposits to banks, cash checks, collect bank statements and have checks certified	4.30

COLLECTIONS AND ADJUSTMENTS	Weight
7. Handle complaints, investigate	7.38
11. Make customer adjustments, smooth out difficulties	8.04
14. Call on clients	3.33
17. Telephone delinquent customers	3.14
20. Make personal calls for credit information	1.83
64. Make personal collection calls	2.99

JUNIOR CLERICAL	Weight
9. Pay bills and bring back receipts	3.07
36. Get mail and stamps at post office	2.74
40. Assist others, general handyman	2.94
41. Take mail, registered mail and parcels to post office	2.46
44. Fold letters, circulars for mailing	1.74
49. Seal, weigh, and stamp mail	2.68
50. Run errands for employer	2.51

61. Take messages, papers, and distribute mail to departments..... 1.94
Score ½ on:
42. Mail out statements and correspondence..... 4.01
60. Address envelopes, bills, etc..... 2.07
66. Take deposits to banks, cash checks, collect bank statements, and have checks certified 4.30

SALES—OFFICE	Weight
4. Check on competitors' prices, compare quotations.....	2.66
15. Dictate letters, reports, etc.....	2.25
22. Attend conference with supervisor.....	3.03
39. Make forms and charts.....	3.65
45. Purchase merchandise, supplies, equipment.....	5.91
48. Make out price sheets.....	2.69
54. Organize work, train, and supervise others.....	4.31
58. Make up sales contracts.....	2.40
71. Keep trade information, credit information, up to date.	2.91
73. Classify orders to size, patterns, salesmen, etc.....	1.45
75. Figure quotations	2.06

SALES—STORE	Weight
1. Wrap up bundles and packages.....	5.50
2. Sell goods over telephone.....	6.10
6. Unpack goods, put away, and keep storeroom in order.....	4.58
8. Give information, quote rates over telephone.....	7.77
10. Put tags and labels on merchandise.....	3.09
13. Deliver orders to customers.....	2.50
18. Dust shelves, put in order.....	2.60
24. Restock shelves and cases.....	4.11
25. Set up displays, window trim, etc.....	4.28
32. Make up and schedule shipments, etc.....	3.98
34. Check and receive incoming supplies, record, etc.....	8.07
43. Arrange display of food stuffs.....	2.61
46. Put up mail and telephone orders, fill orders to be shipped.....	4.17
47. Letter signs for stock display.....	1.38
53. Check on quality of goods, examine for defects.....	7.53
59. Sweep floors, empty waste baskets, clean up, etc.....	2.04
62. Dismantle window displays.....	1.49
67. Take inventories	6.26
72. Wait on customers, sell over the counter.....	7.95

STENOGRAPHIC-FILING	Weight
19. Answer telephone	5.08
28. Take dictation and transcribe.....	4.92
30. Code and type telegrams.....	1.51
51. Type letters, orders, forms, etc.....	3.95
56. File orders, letters, bills, reports, trade information.....	4.67
65. Look up information files, library, etc.....	7.46
Score ½ on:	
38. Type financial statements	2.14
42. Mail out statements and correspondence.....	4.01
60. Address envelopes, bills, etc.....	2.07

It may be observed that these weights run from 1.25 to 8.07 with the greater part of them falling between values of 2. and 5. Since only one value exceeded 7.99, unit weights of from 1 to 7 were used (determined by value without regard to decimal) in scoring each item in respect to the pattern in which it belonged.

Final Interest Test

We have now arrived at a list of 75 specific job activities which are common to initially available business jobs and which fall within specific patterns; we have determined, too, their relative significance within these patterns. An individual's reaction, therefore, as to the extent to which he feels that he would like or dislike these activities, can now be evaluated in terms of specific beginning business positions. Before setting up these items in terms of responses, however, we must consider how they shall be listed. To merely list them in the order in which they appear under their pattern headings would obviously condition responses which seemed to go together. A random order seems desirable. Accordingly the listing by patterns is now numbered from 1 to 75, and these numbers converted into a random order by Fisher and Yates' Table of Random Numbers.[11]

On the final form[12] the questions appear directly on an I.B.M. answer sheet. Instructions are also printed on this sheet and space provided for name and pertinent information concerning the individual taking the test. The instructions are reproduced here but the items have appeared earlier, though in different order.

Instructions

This questionnaire is designed to indicate how you feel about those specific job-activities which characterize initial business positions. You are to indicate your answer by blackening the space between the proper pair of dotted lines.

The first three columns are headed L I D so that you may record your response as *like, indifferent,* or *dislike.* If you think that you would like to perform the job-activity indicated as a part of your first business job, record your

[11] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research,* (London, Oliver and Boyd, 1938).
[12] Published by Science Research Associates.

response under L, if you feel uncertain or indifferent, under I, if you feel that you would dislike this activity, under D. Omit no items.

When you have completed this, go over the items again carefully and indicate in the X column the *five* which you would *most like to do*. There is no time limit, but you should work fairly rapidly as it is your first impression which is important.

Scoring

The L I D responses are, of course, familiar to all. Here the "L" response is scored with the weight previously determined within its respective pattern. The "I" response is scored as one-half that amount with all fractions dropped, with the "D" response scored as zero. The "X" response is new to previous practice in this type of test, and allows an opportunity for an individual to distinguish a little more carefully between those job-activities which he feels he would like as part of his initial job. It serves, too, as an additional aid to the counselor, apart from the resultant pattern scores. The fact that the individual selects five out of 75 items as *most* to be desired should result in additional weights in respect to such items. We may logically expect these selections to be made in respect to items in which an "L" response has been recorded, and an additional weight equal to one-half the weight of the "L" response should result in some refinement in scoring without being regarded as excessive. This "X" response can be scored without additional runs through the machine.

The test may also be scored by hand. Special hand-scoring folders are provided which hold the sheet in position for the proper registration. The norms appear directly on this folder which is so cut that the score has to be recorded in the right place each time. For machine-scoring two keys are provided for each pattern. By picking up a number of "contrasts" each time only two runs are necessary to pick up positive weights of from one to seven; thus machine and hand-scored norms

are identical. Since it is the same sheet in each instance, test conditions are also identical.

Directions for Administering

Some students will complete this test in twelve minutes; the majority, however, finish in approximately fifteen minutes, and none takes over twenty minutes. The subjects' first run through in respect to the L I D responses is done rapidly and without hesitation, and this observation leads us to believe that there is very little doubt in their minds as to how they feel about these particular job-activities. Some hesitance, however, is seen in selecting the five job-activities most to be preferred, which is equally desirable, as it indicates a tendency to weigh them carefully.

Intercorrelations, Reliability and Validity

TABLE 8

	Acctg.	Coll. & Adj.	Jr. Cler.	Sales Office	Sales Strs.	Sten. Filing
Accounting	.92	—.25	.08	.22	.10	.22
Coll. & Adj.		.73	—.13	.27	.00	—.05
Jr. Clerical			.78	.01	.65	.41
Sales—Office				.78	.26	.07
Sales—Store					.77	.31
Sten.-Filing						.80

In Table 8 we present the intercorrelations of these patterns with their reliability coefficients as diagonal values. With the exception of the relationship between the Junior Clerical and Sales-Store patterns these coefficients are low enough to indicate satisfactory independence of these patterns. The correlation of .65 between the two mentioned, however, is too high to be disregarded. Further, such raw coefficients understate the actual relationships involved. To eliminate as much of the chance factors as possible we apply the formula[13] $\frac{r_{ab}}{\sqrt{r_a}\ \sqrt{r_b}}$ to correct for attenuation. The corrected value of .83 clearly indicates that these patterns do not act independently. A study of the relation of these two patterns

[13] C. Spearman, *American Journal of Psychology*, XV (1904), p. 271.

to the other four provides still further evidence that these two patterns should be combined.

	Acctg.	Coll. & Adj.	Sales Office	Sten. Filing
Junior Clerical	.08	—.13	.01	.41
Sales—Store	.10	.00	.26	.31

Combining these two patterns would also serve to step-up the reliability, as this coefficient is materially affected by the range of scores made on a test.

TABLE 9

	Acctg.	Coll. & Adj.	Sales Office	Sales Strs.	Sten. Filing
Accounting	.92	—.25	.22	.13	.22
Coll. & Adj.		.73	.27	.06	—.05
Sales—Office			.78	.13	.07
Sales—Store				.86	.37
Sten.-Filing					.80

Table 9 represents the revised matrix on the basis of five patterns with the diagonal values representing the reliability coefficients as in the preceding table. The highest relationship now observable in this revised matrix is that of .37 between Sales-Store and Stenographic-Filing, and even this is satisfactorily low. Considered critically, it becomes .44 when corrected for attenuation; squaring it, it becomes .19, and gives us an idea of the variance which can be accounted for in this relationship. More significant in our present consideration is the amount of variance not accounted for, given by $1 - r,^2$ — in this case .81.

It will be noticed that the combination of the two patterns which were not independent now has a reliability of .86. The lowest reliability coefficient occurs in the Collections and Adjustments pattern with a .73, having only six items, and the highest reliability coefficient in Accounting with .92, a pattern comprising 23 items.

As has been pointed out, the size of a reliability coefficient is partially determined by the number of items, particularly when computed by the method of split-halves[14], and it might be argued that more items should have been retained in setting

[14] Karl J. Holzinger. "Note on the Use of Spearman's Prophecy Formula for Reliability," *Journal of Educational Psychology*, XIV (1923), 301-305.

up these tests. The retention of such items, however, would have been at the expense of the validity of the test, since the test was constructed with this objective initially in mind. This approach deviates materially from the common practice in the construction of interest test questionnaires, which are usually built from items empirically determined, and then scrutinized to determine their validity. Items, of course, are usually rated before inclusion in regard to relevance and importance, and the ultimate validity determined by the analysis of scores obtained by persons successfully engaged in such occupations, or still more feebly, by the opinion of experts as to internal consistency. The validity of these items was determined by a cluster analysis, and only those which acted as pattern-determiners were used in the test.

Immediately following are norms for these final five patterns in tabular form. They are given as standard scores and normalized scores with $M = 50$ and $\sigma = 10$, and are based on 304 freshmen at Boston University, College of Business Administration. Additional norms are available but are not necessary in the clinical use of the instrument. What is important is the "level of significance" which is here considered as the upper half of the range.

TABLE 10

Percentile	Standard Scores	Raw Scores				
		Acctg.	Coll. & Adj.	Sales—Office	Sales—Store	Stenog.-Filing
100					100	32
99	2.32			36	91	29
98	2.05	74		34	84	27
95	1.64	68	30	32	77	24
90	1.28	62	27	30	69	22
80	.84	56	23	27	62	19
70	.52	50	21	25	56	17
60	.25	46	19	23	51	15
50	.00	42	17	22	46	13
40	— .25	38	14	21	41	11
30	— .52	34	12	19	35	9
20	— .84	28	10	17	29	7
10	—1.28	22	6	14	22	4
5	—1.64	16	3	12	14	2
2	—2.05	10		9	7	
1	—2.32	6		7	2	

Present Use of This Instrument

This test is now being used for several purposes. First, it assists the freshmen of the College of Business Administration in the selection of academic majors. Interest in the particular job-activities involved in the occupational areas for which such majors train is by far the most important single consideration in making such selection. True, the probable "risk" of pursuing such training in terms of special abilities and personality traits still remains to be evaluated. Many instruments, however, are available for diagnostic purposes, and each entering freshman now regularly takes a battery of such tests upon admission.

Second, it was used in the Evening Division of the College of Business Administration of Boston University where the immediate pursuit of an initially available job is economically imperative. The "X" response in particular has been found helpful to the counselor when the student comes to his desk, and the whole approach seems to better motivate the individual's interest in the specific job-activities of beginning positions which he has been considering tentatively.

For commercial students at the senior high school level who are looking for jobs the results should have the same significance as for the evening student just mentioned; they, too, need indicators for beginning positions.

The test is also being used experimentally at the 9th grade level within the commercial curriculum to distinguish between bookkeeping, sales, and stenographic interests where such help is badly needed. In situations such as found throughout New York State where a course called "Introduction to Business" is given to all commercial ninth-graders, no problem as to terminology of test items is met; but where no such orientation course is given, several terms need discussion and explanation. The maximum usefulness of this direct testing technique rests on an extension of realistic work experience provided for in the curriculum.

In out-of-school situations this test likewise has several applications. Social agencies, faced with a need for specific

counseling previous to job hunting, find it extremely useful as a direct pointer towards specific beginning jobs. In the Guidance Department at the Boston YMCA, for example, where an excellent and extensive job in counseling is being done, this test is basic to all batteries. All members of Darling's "Job Hunters" group at Boston take this test since a large part of them are interested in beginning business jobs. Employment managers likewise in several situations are using this test for beginning office workers where the present scarcity of beginners places more emphasis on allocation than merely selection.

Implications Arising From This Study

The usefulness of this test is obviously limited within the range of an already generally determined interest in the field of business. What is still needed is a general interest test of this functional type which would allocate student interest into general areas. This accomplished, areas needing a more detailed diagnosis would be indicated. An extension of the horizontal testing range for initially available positions is equally feasible in other areas. As has been previously indicated, maximum effectiveness of this positive approach to interest measurement depends upon the progressive development of the secondary school curriculum toward increased emphasis on realistic bits of work experience. Progressive educators predict that within a few years 25 per cent of the ninth-grade work may be of this type, increasing until 75 per cent of the twelfth-grade level curriculum will be composed of such work experiences. Such a program will go far toward bridging the present gap between school and job placement.

It should be similarly noted that a vertical extension of this testing technique is equally possible. Actual experience on a job results in a refinement of interests in the field and should be measurable as a pattern indicative of progressive specificity. Many quite different jobs may be indicated from the development of more specific interests within a single initial pattern, and assistance in the selection of secondary level jobs could thus be provided. The technique for constructing such interest tests is now available.

MEASUREMENT IN RURAL HOUSING
A PRELIMINARY REPORT[1]

CHARLES I. MOSIER
Social Security Board

THE SLOAN PROJECT in Housing Education at the University of Florida is an attempt to investigate the broad problem of the extent to which educational materials introduced through the school may exert an influence on the social, cultural and economic life of the community as a whole. Specifically, the problem investigated is the effect on the housing status of the community, and on the housing attitudes of its members, of a broad program of education in all aspects of housing introduced through the schools. The general outlines of the program involved the utilization of six white school communities of Florida, three experimental and three control. The plan of the experiment involves a determination of the present status of all six communities in those attributes which might be affected by housing education, the introduction of housing material into the regular curriculum in the schools of three experimental communities, and subsequent tests to determine the change in housing status, attitudes, and school achievement which can be attributed directly to the experimental program.

The present report is concerned only with a summary of the initial measurement aspects of the experiment.[2] All measurements have been made in both experimental and control communities at the beginning of the experimental program to

[1] The study reported here is being carried on at the University of Florida under the auspices of the Sloan Foundation in Applied Economics.

[2] A full report of the initial measurement program has been prepared and is on file at the University of Florida Library. C. I. Mosier, *Measurement in the Field of Rural Housing* (1941).

establish a base line from which to measure change, and they will be repeated at subsequent intervals throughout the course of the experiment as well as at its close. Certain of the measurements might well be made again several years after the termination of the experimental work of education in order to determine the stability of the results and to provide a measurement of those changes occurring after relatively long latent periods. As an instance of the latter, we might hypothesize that one of the outcomes of the experiment would be that those pupils who had been exposed to the experimental program of housing education would, on leaving school and establishing homes of their own, secure more adequate housing than those who had not. The observation of the full impact of such effects could take place only twelve to fifteen years after the beginning of the experiment. Twelve years would be required for the pupil to progress normally from the first grade to high-school graduation and thus receive the fullest benefit of the educational program. At least three more years would elapse before any sizable proportion of the subjects would have married and become sufficiently established economically to provide themselves with homes which could be considered indicative of their ultimate housing status. It is not proposed that such an extended period of observation is essential to the program, but attention should be called to the long-range character of certain of its effects.

The measurements which have been undertaken can be divided into four major areas: housing adequacy, housing attitudes and insight, housing information of pupils, and academic achievement. For the measurement of housing adequacy we have developed a *Housing Inventory*[3] describing the objective condition of the house as observed by a trained field-worker, yielding a *Housing Index*—a composite score for the house obtained by weighting and combining these observations to provide for each house a single score. The inventory records

[3] For a detailed report on the development of the *Housing Inventory*, see C. I. Mosier, *Measurement of Rural Housing Status*, in preparation.

obtained by interview have been supplemented, in a large proportion of the cases, by photographic records of the houses studied.

It is conceivable that the program of education might produce a real change in attitude toward housing, in insight into the present inadequacies where they exist,[4] and in motivation toward better housing conditions, and yet there might be no externally observable improvement because of the pressure of economic circumstances. Because of this, an attempt has been made to measure the extent of such effects by a separate evaluation of the answers to certain of the inventory questions. Plans have been made for a more direct measurement of attitudes and for the development of tests measuring achievement in the acquisition of information in the field of housing, but these plans have not yet been fulfilled.

It is important to know whether the introduction of housing materials into the curriculum has been at the expense of training in the fundamental academic achievements, or whether it has resulted in more effective learning of the skills of reading, arithmetic and language through a use of material of more immediate interest and appeal than that contained in the customary curriculum. As the initial phase in the investigation of this problem, a program consisting of four achievement tests and an "intelligence" test has been administered in grades 4-12 in all schools. The results of this initial testing and the subsequent retesting which is contemplated will provide valuable information on this question and on others which may arise.

Before beginning a summary of the initial results in each of the four areas of measurement, certain further general statements should be made concerning the communities investigated and the nature of the sample involved. While the detailed description of the several communities can best be presented in the light of the results of the initial surveys, cer-

[4] In answer to the Inventory question, "What changes or improvements do you think are needed", the occupant of one extremely dilapidated shanty said, "Well, if I could get hold of some cardboard boxes to tack up inside to cover the chinks, I reckon I'd have a right tight little place."

tain broad generalizations can be made which will facilitate the understanding of these results. For the purposes of this study, a community is defined as consisting of all those families, and only those families, which send at least one child to a specific school being studied. This automatically restricts the population to white families. Such a definition necessarily involves a somewhat different use of the term "community" from that ordinarily envisaged. It excludes all those families, maintaining their identity as family units, which have been established so recently that there is no child of school age. It excludes any families who have children of school age, but who, for one reason or another, do not send their children to the school in question. In one of the communities the field-workers reported this situation: "Are you going to see the Joneses down the road? They've got a flock of kids, but the kids don't go to school because they ain't got no clothes." The extent to which this, and other comparable situations, prevail is, at present at least, unknown, but it does exist and influences the sample studied, since the group under discussion will consist of the poorest families in the area. In certain of the communities the definition of the population imposes another restriction in the outlying districts. As the distance from school becomes great, the children of grade-school age go to the local school, so that only those families with at least one child of high-school age are included. The inclusion of a family in the "community" depends, then, not only on the age-distribution of the children (and hence the length of time the family has been established), but on geographical location as well. These selective factors will, inevitably, limit the extent to which the results of this study can be generalized to the community-as-a-whole, since the population studied is limited to those families sending at least one child to a participating school.

The unit of the investigation, when it is not the individual pupil in the school, is the dwelling group. All persons living within the same dwelling-unit (house or separate apartment) are considered to constitute a "family unit." A total of 745

HOUSES AT SELECTED LEVELS OF HOUSING INDEX VALUE

Score 17

Score 20

Score 23

Score 26

Score 29

Score 32

HOUSES AT SELECTED LEVELS OF HOUSING INDEX VALUE

Score 35

Score 38

Score 41

Score 44

Score 47

Score 50

families was studied, with test records of 1028 children of school-grade above the fourth. There are records of 522 children in grades 1-3, but these records are incomplete and do not represent the total number of children in those grades. The population was defined as of the date of interviewing (October and November, 1940), and families moving into the community after that date were not considered.

All of the primary data and much of the derived data have been recorded on electric accounting equipment cards, where they are readily available for further research.

The evaluation of the housing status of the communities was by means of a *Housing Inventory* specifically designed to measure housing adequacy in rural areas. In addition to identification data the *Inventory* recorded the responses to 85 items of the type:

Fireplace and chimney—state of repair? 1. no fireplace. 2. poor—masonry cracked, mortar crumbling, many loose bricks. 3. fair—masonry discolored, occasional loose bricks. 4. good—no obvious repairs needed.

Kind of cookstove (if more than one, mark the highest number). 1. open fireplace. 2. makeshift stove—sheet-tin arrangements, etc. 3. small wood stove. 4. large wood range. 5. electric. 6. gas, bottled or city.

"How many rooms, not counting closets, porches (or the bathroom) are there?" (check answer with your own observation). 1. one. 2. two. 3. three. 4. four. 5. five, 6. six. 7. seven. 8. eight. 9. nine. 10. ten. 11. eleven. 12. twelve or more—specify.

With the exception of answers to certain specified questions the observations recorded were those of a trained interviewer, not of the occupant.

The criterion which such an instrument should fulfill was established, available sources searched for suggestions as to possible items, and a preliminary edition prepared. This preliminary edition was subjected to the editorial scrutiny of research workers in the field of rural welfare, revised, and subjected to a searching field-test in which two Inventories were completed for each house under conditions comparable to those of the main study. The *Inventory* was revised again in the light of the field experience, and prepared in final form.

Three interviewers meeting a set of pre-established qualifications were used. They were given intensive training in the use of the *Inventory* under actual field conditions and in establishing common standards of evaluation. A booklet of *Instructions to Raters* was prepared for their further guidance. All coding of the data and computations were performed separately from the interviewing, and a routine of computation established. An occupational code adapted to the needs of the particular problem has been devised, and applied to the classification of the heads of the families studied.

The data from the *Inventory* have been recorded on punch cards. A procedure has been devised to weight the responses to the individual items in such a way as to provide the maximum accuracy of measurement.[5] The weighted responses have been combined into a single composite *Housing Index* measuring the adequacy of each house as a dwelling place. The reliability of this *Index* has been estimated in several ways. In a selected area separate from the six communities, but typical of them, 50 houses were inventoried twice by a different interviewer, and with a time interval ranging from one to five weeks in order to estimate the errors due to the interviewer and the time of interviewing. The consistency of the scores on this test-retest survey was exceptionally high ($r = .96$), and no systematic difference was found between the two interviewers studied. Estimates of reliability by the split-halves technique of determining reliability and by the method of rational equivalence both yielded reliability coefficients in excess of .97. The accuracy of measurement reflected by these values can be seen from the following considerations: 32 per cent of the houses received their true scores, and no house received a score in error by as much as four points out of a range of 35 points.

The *Index* was validated in part by the criterion of internal consistency. The extent to which the weighting procedure automatically transformed certain first approximation

[5] A description of the statistical technique of weighting the item-responses is being prepared for early publication.

weights which were not in accord with *a priori* considerations into weights which agreed closely with reasonable values was adduced as further important evidence of validity. A high degree of internal consistency among the *Inventory* items was revealed—houses good in one respect tended to be good in all respects, and conversely. As further validation the photographs of one hundred houses were scaled for adequacy by the psychometric method of equal-appearing intervals and the correlation between these scale values and the *Housing Index* compared. The relationship was more than satisfactorily high, ($r = .81$), but not high enough to justify substituting photographs for *Inventory* ratings. The meaning of individual *Index* scores is further made graphic by the presentation of photographs of actual houses for selected values of *Index* score (shown in the illustration Figure 1). The typical housing conditions at several score levels have been described and are presented in detail elsewhere as aids in the calibration of the *Index*.[6]

Some of the more significant descriptive findings of the housing survey are presented here. The median family consists of two adults, two children over twelve and one child under twelve years of age. Sixty-two per cent of the families own their own homes, 18 per cent rent, and 20 per cent are classified as share-croppers, squatters, or rent-free tenants. Fifty-nine per cent of the families gave farming as their only occupation; 12 per cent stated that they were on relief, or gave "public work" as their occupation.

Twenty-two per cent of the houses have only three rooms, or fewer, but 17 per cent have rooms closed off and not used, unless for storage. Thirty per cent have less than one room for each adult or equivalent; 45 per cent have no separate living room. Thirty-eight per cent used auxiliary bedrooms (rooms used during the day for some other purpose); 27 per cent sleep with two persons or more in every bed, and in at

[6] C. I. Mosier, *Measurement of Rural Housing Status, loc. cit.*

least 14 per cent of the houses, sex privacy in sleeping arrangements cannot be maintained.

Forty per cent have inside walls completely unceiled, with the studding showing. Eight per cent have no decoration of the inside walls whatever, but only three per cent utilize handicraft decorations or native materials. Fifteen per cent have no pantry or storage space in the kitchen, or only poor makeshifts; 80 per cent have no kitchen sink whatever, and 52 per cent have no refrigeration whatever; 72 per cent have no electricity. Eleven per cent of the families must carry their water more than a hundred feet, and another 61 per cent have only outside hand pumps or wells. Fifteen per cent have yards littered with garbage and refuse; 16 per cent have no toilet facilities whatever, and another 70 per cent have no better than an open surface privy. The prevalence of hookworm, typhoid, and dysentery is not surprising.

Sixty-three per cent of the houses have chimneys which were judged to constitute some degree of fire hazard. Twenty-seven per cent of the houses have unglazed windows (wooden shutters only), and another 16 per cent have more than three broken panes. Only 41 per cent have all outside openings screened and in good repair to serve as protection against malaria or typhoid. The roof needs some repairing in 47 per cent of the houses, and in 6 per cent one can see daylight through it; 24 per cent show visible evidence of termite damage—only 2 per cent are termite-proofed—and 52 per cent show some degree of damage from dry-rot; 13 per cent have their foundations sagging, rotted, and crumbling. Sixty-three per cent showed no evidence of having been painted at any time; 37 per cent have no shrubs around the house and 28 per cent have no flowers; 58 per cent are lacking even bordered and sand-surfaced walks, while less than one per cent used pine-straw to surface the walks and drives.

In spite of these objective conditions, 44 per cent of the occupants mentioned no more than one aspect of the house needing repair; only 12 per cent actually planned repairs, and

the average family *says* that, if they won a hundred-dollar prize or found that sum, they would spend $57.00 on the house.

The results of the survey have been tabulated for each community, and for the experimental and control groups, both in terms of the percentage response frequency for each item-response, and of the frequency distributions of the *Housing Index*.[7] The differences between the experimental and control groups were systematically examined. Differences in housing status between the individual communities are very great—the best house in the poorest community is not as good as the average house in the best community. Differences between the experimental and control groups are small, the control group showing a slight superiority.

Photographic records have been obtained for 517 of the houses studied. These photographic records were obtained under standardized conditions, so that they may be repeated at a later date. Map records showing the location of each house in each community have been prepared and are on file. The possibility of analyzing these data to show relations with geographic factors is considered.

Achievement tests in reading, arithmetic, language, science, and mental maturity were given to 1028 pupils in all grades above the third. The results of this testing program have been analyzed by community and for the experimental and control groups. There was no discernible difference in the relative school achievement for the experimental and control groups, although there was considerable variation among the schools themselves. All schools were, on the average, markedly retarded in achievement as compared with the chronological age or the grade placement of the students. This mean retardation was from one and one-half to nearly three years, most marked in the higher grades, of course, and greater in science achievement than in any of the other fields. When,

[7] These data are presented in full in *Measurement in the Field of Rural Housing, loc. cit.*

however, achievement is compared, not with chronological age or grade-placement, but with mental age, this apparent retardation disappeared, so that it can be said that the schools are educating the pupils to the limits of their mental capacities—assuming that the intelligence test does measure mental capacity.

The relationship between school achievement and housing conditions has been investigated. In spite of the reasons to expect that achievement would be related to the conditions of the home, the results do not bear out this expectation. The *Housing Index* showed correlations which were positive, but very low—ranging from .12 to .31—with the various measures of school achievement. The most significant relation was between *Index* and grade placement, indicating that children in the higher grades tend to come from superior homes.

Certain items of the *Inventory* were designed to measure, not the adequacy of the house, but the attitude of the family toward housing problems—insight into the housing condition, and motivation to better those conditions. When these items were studied by multiple factor analysis, the existence of a single factor of housing attitude, independent of housing adequacy, was convincingly demonstrated. This attitude variable is measured by the items dealing with willingness to spend money on the house, with ownership, with the number of repairs wanted by the occupant, with the difference between repairs needed and repairs wanted, and with whether or not repairs were planned. A method of measurement of the strength of this attitude in each family has been devised and is being applied to the individual families.

Detailed plans for a more direct attack on the problem of measuring attitudes by means of a specially designed attitude scale have been prepared. The development of this scale and its application to the families studied is a project which, it is hoped, will be undertaken at the earliest possible opportunity.

One of the contributions of this study, apart from the development of a *Housing Inventory* and its standardization,

has been the accumulation of data for subsequent analysis in connection with specific problems which will serve as the base line from which the effects of the experimental curriculum can be judged. These data—coded answers to 93 items in the *Housing Inventory* for each of 715 houses, a measure of the adequacy of each house, a record of its location and photographs for 517 of the houses, measures of school achievement for each of the 1028 children in grades 4-12 of the six schools, and summaries of the frequency and percentage frequency of each of the 685 item-responses of the *Housing Inventory* for each of the six communities and for the experimental, control and total groups—have been recorded and filed on International Business Machines punch cards, and detailed indices to this information are presented in the detailed report.[8] How valuable it is will depend on the extent to which these data are used to provide a knowledge of the factors affecting rural housing, whether those factors be educational, sociological, psychological, or geographical.[9]

The principal results of the initial measurement program can be summarized as follows:

1. A *Housing Inventory* has been prepared and applied to the 715 homes of children in six rural Florida schools.

2. A technique for weighting these responses to the *Inventory* to yield a measure of housing adequacy, the *Housing Index,* has been developed.

3. A *Housing Index,* using the weights obtained, has been carefully standardized, using a group in addition to that on which the weights were developed. The reliability coefficient by test-retest with different interviewers was .96 and by internal consistency was .97. The *Index* has been validated by expert opinion, by internal consistency, and by comparison with psychophysical scaled judgments of adequacy based on photographs. The correlation coefficient between *Index* and scale values from judgments of photographs was .81. The

[8] *Loc. cit.*

[9] These data will be made available to any interested workers engaged in problems toward the solution of which these data might contribute.

meanings of the various *Index* scores in the complete report have been interpreted by describing the typical houses and presenting photographs at various score levels.

4. Standardized intelligence and achievement tests in six fields were administered to all the children of the schools. The correlation of children's school achievement with home conditions as measured by the *Housing Index* was low for all measures of achievement, with coefficients ranging from .12 to .31.

5. By utilizing the answers of occupants to selected items in the *Housing Inventory*, a factor of housing attitude has been isolated, and initial attempts to measure it have been made.

PROCEDURES FOR HANDLING TESTS AND EXAMINATIONS

JOHN V. McQUITTY
University of Florida

THE BOARD OF UNIVERSITY Examiners of the University of Florida conducts a program of testing somewhat different from that ordinarily performed by examining boards. Throughout the school year it offers a regularly scheduled series of progress tests in the basic courses in the General College. These tests are given in addition to the comprehensive examinations which are given at the completion of each course. The Board integrates the International Test Scoring Machine and punch card tabulating equipment to enable it to report results of tests promptly and adequately to all persons concerned. Also, the Board uses punch cards in building its library of test items. It is the purpose of this paper both to discuss the general work of the Board and to give the operations in detail, with special emphasis on the use of punch cards.

At the University of Florida all of the freshmen and sophomores enroll in the General College for their first two years' work. The Board of University Examiners was created in 1935 along with the establishment of the General College. The Board was charged with handling the admissions to the University and with the examinations given in the comprehensive courses which were to be offered in that college. At present the enrollment therein is about two thousand. The college examining activities of the Board come under two heads: comprehensive examinations given at the completion of the courses, and progress tests given at regular intervals of from two weeks to a month in each of these courses. The comprehensive examinations are six hours in length for two-

semester courses, and three for semester courses. The results of these examinations form the sole basis for the assignment of the student's final grade. The progress tests are similar to the comprehensives except that they cover smaller areas of the course and are usually only one hour in length. These tests are given to indicate to the student, his instructor, his parents, and the University officials how each student is doing. Even though some of these tests are given as early as eight months prior to the comprehensive examinations, the coefficients of correlation between results on progress tests and comprehensive examinations range from .65 to .83. Thus the importance of the progress tests as indicators of probable success on the comprehensive is demonstrated.

When the progress testing program was first instituted, both students and faculty were somewhat skeptical of the value of progress tests since their results were not counted when the final grades were assigned. For one thing, the practice of *not* averaging in test results at the end of the course was a new and radically different procedure and hence subject to view with considerable alarm. Now that several years have shown that the progress tests are just as important whether or not they are included in the final grade, the question of their value is no longer raised, but their usefulness is taken as a matter of course. There are two definite reasons for basing the grade entirely on the final comprehensive: 1. The grades are then assigned on the basis of how well the examinee knows the course as a whole, since piecemeal learning of the material is not enough to insure success on the examination. 2. Under such a practice the progress test results become sign posts along the way which indicate how the student and instructor are working together so that the former may achieve success on the comprehensive, but the student who compensates for an inadequate preparation and a consequent poor showing on the early tests by ultimate mastery is not penalized for his early failure.

Examinees are permitted to keep the progress test booklets, and these constitute an excellent source of material for

review. Also, the examinees' answer sheets are returned to them. Thus the student not only has a record of his raw score and his percentile rank, but also a record of the answers which he gave to the questions. By studying these answers in relation to the key of correct answers which is returned with the answer sheet, the student can make a detailed study showing which items were missed and which were answered correctly.

In addition to the tests and examinations already discussed, the University sponsors each spring a state-wide twelfth-grade high-school testing program, the results of which are used by the University as entering placement tests. The General College handles the announcements of the program and the distribution and receipt of the test materials.

Separation of Instructing and Examining

In theory there is complete separation of the teaching and examining functions. The examining and the issuing of final grades are both done by the Board of Examiners. However, in actual practice there is the closest cooperation between the instructional staffs and the Examiners. In most of the courses the construction of the test items is done by persons engaged in teaching those courses. These items are then subjected to critical review by the Examiners, and any test items composed by the Examiners are reviewed by members of the instructional staff. In all cases the test or examination has the approval of both the instructional staff and the Examiners. All tests are printed, given, and scored by the Examiners. When it comes time to set the grades—i.e., determine the raw-score division points for the passing grades A, B, C, D, and the failing grade E—the members of the instructional staff cooperate again. The members of the staff and a representative of the Examiners hold a meeting scheduled for this purpose. In this task, use is made of all pertinent objective information about those students making a given raw score; the results of the entrance examinations and of the progress tests throughout the year as well as the characteristic responses to those

test items considered most crucial are all considered for students at the critical values dividing letter-grade equivalents. Also at hand are the distribution of raw scores on the examination and the corresponding percentile ranks. The anonymity of the individual student is preserved throughout this procedure. Again it must be made clear that all these data are used as aids in determining where the grades should come on the distribution of raw scores. In no sense are any of the data "averaged-in" when the grade is assigned. Once it is decided, for example, that scores of 400 or above are to receive A's, everyone in that category—but no one else—receives an A, and so on for the other grades. There is no "grading on the curve" in the sense that a predetermined distribution of grades is followed. This is shown by the fact that even in a course taken by as many as 700 persons the percentages of A's has varied from 5 to 11 and the percentage of failures from 12 to 21. Since 1935 the Board of Examiners has assigned 42,214 final grades with the following distribution:

TABLE 1

DISTRIBUTION OF GRADES FOR COMPREHENSIVE EXAMINATIONS WINTER, 1936 THROUGH SUMMER, 1941

Per Cent for Each Grade*					Total	Per Cent	
A	B	C	D	E	Examined	Absent	Total
8.42	16.72	37.35	21.85	15.66	41,112	2.68	42,214

*Based on number examined.

Office Routine

In discussing the routine for handling the test results and the test items, special emphasis will be given to those procedures which may not be widely known. It is recognized that practices for the processing of examinations will vary according to the use to be made of the test materials and with the mechanical equipment available. The Board of Examiners makes use of the following mechanical equipment supplied by the International Business Machines Corporation: test scoring machine with graphic item counter, alphabetical printing punch,

high speed reproducer, interpreter, collator, and alphanumeric tabulator with 25 alphabetical and 30 numeric type bars. The routine of handling the progress tests is affected by the facts that all answer sheets are to be returned to the students, and that the sheets carry the raw scores and the percentile ranks. In the case of the comprehensive examinations the answer sheets are retained by the Examiners, the examinee receiving nothing but his letter grade. Also, the need for prompt reporting of results is particularly great in the case of progress tests because it is felt that the results are more helpful to the students if received while interest is still high. Hence, progress tests are usually given on a Saturday morning and the results returned to the instructors, administrative offices, and the students the following Monday, even though as many as 1000 students are given two tests each. In the case of the comprehensive examinations there is an equally great need for prompt work because all examining for the year's work for the freshmen and sophomores must be accomplished and final grades submitted within a period of two weeks. It would be impossible to do all this on the limited budget of the Examiners without making full use of the punch-card system. The integration of punch-card methods with the test scoring machine, however, permits a very small staff to handle a large volume of tests in a short time and to make the results available in a variety of forms to meet the varying needs of student, instructor, department head, and administrator. The use of punch cards will be discussed under the four following heads: (1) placement tests; (2) progress tests; (3) comprehensive examinations; and (4) library of used test items.

A. *Punch Cards and Placement Test Results.*—In the spring of 1941 the following tests were used in the high-school testing program:

1. The Henmon-Nelson Test of Mental Ability, Form B
2. Cooperative English Test, Effectiveness of Expression, Lower Level, Form Q

3-5. Cooperative Achievement Tests, Form QR, in
   3. Social Studies
   4. Natural Sciences
   5. Mathematics
6. Cooperative French Test
7. Cooperative Latin Test
8. Cooperative Spanish Test

All of these tests are given with separate answer sheets which are machine-scored by the Board of Examiners. All of the test results are punched into tabulating cards, and the following steps are employed in the procedure:

1a. Answer sheets are handled by schools, and when the sheets are received, they are separated by tests and alphabetized. Some visual check such as a corner-cut or a punched hole will aid in the checking of the homogeneity of a pile of answer sheets.

2a. Name cards are punched for each examinee. Each card carries a code number indicating the high school. Sex is indicated by using F for female and M for male. A heading card with a characteristic control punch is made for each high school.

3a. The name cards are listed on the tabulator in alphabetic order by high schools on a prepared form which carries eight columns, one for the raw scores for each test.

4a. The answer sheets which have already been separated according to tests and alphabetized are checked against the list described in 3a, to make certain that the answer sheets are in the same order as the names on the list. In the case of persons who took one but not all of the tests, the areas where results for the missing tests would be recorded are marked. The practice of having the answer sheets and names in identical order, with absentees designated, facilitates the recording of the machine scores.

5a. The answer sheets are scored on the test-scoring machine and the raw scores recorded in the appropriate rectangle on the name lists. Usually two persons are used, a machine operator and a recorder. The operator gives the scores orally to the recorder, who enters them in the proper place.

6a. The answer sheets are scored again and the raw scores checked against those recorded in number 5a.

7a. The checked raw scores are punched into the name cards. The name cards and the lists carrying the raw scores are in the same order.

8a. The name cards are listed on the alphanumeric tabulator to show name, sex, and raw scores. A comparing control is maintained on high-school code number.

9a. The lists in number 8a are checked orally against the lists on which the raw scores are written.

Note: From now on all operations are entirely mechanical.

10a. As soon as all of the scoring and punching has been done, the distributions of raw scores are made by sorting the cards in order by raw scores and running them through the tabulator with a comparing control on raw scores if an interval of one is desired (if any other interval is desired, it is necessary to use interval heading cards to establish the control breaks); and progressive totaling is used to secure the cumulative frequencies.

11a. Percentile ranks are computed for the distributions obtained in number 10a. The percentile ranks are punched into heading cards which carry raw scores and corresponding percentile ranks. The corner-cut on the heading cards should differ from that on the score cards.

12a. These percentile rank cards, which must carry an appropriate control, can then be collated into the raw-score detail cards. (Steps 10a, 11a, 12a, and 13a should all be done for one test before anything is done for another test, if the sorting is to be kept to a minimum.)

13a. By using the high speed reproducer the proper percentile ranks are punched from the heading cards into the detail cards, using a control punch to clear the punch magnets at the proper time.

14a. After all of the gang punching has been done (all gang punching should be sight checked and the detail cards checked on the collator for proper sequence order), all percentile ranks can be interpreted at one time.

15a. The detail cards, which now carry percentile ranks, are sorted alphabetically on three letters, then sorted by high schools using the high-school code. Then the detail cards are checked by hand to insure correct alphabetical order.

16a. Lists are run for each high school, showing all percentile ranks for each examinee. Under some conditions it may be desirable to run master alphabetic lists before the detail cards are sorted by high schools.

17a. A master list of the results for all examinees in alphabetic order is prepared on the alphanumeric tabulator. Usually this list is prepared on a stencil or duplicator paper, so that a large number of copies can be made.

B. *Punch Cards and Progress Test Results.*—The nature and use of progress tests has already been discussed. Punch cards are used here as an aid in making the results available quickly to the instructors and administrative officers. The steps in preparing the cards and using them are as follows:

1b. As soon as registration is complete, the class cards in the Registrar's Office are duplicated for each course in which progress tests are to be given. The information picked up from the Registrar's card is:
(1) Student number
(2) Student name
(3) Course and section

2b. The cards prepared in number 1b are alphabetized for three letters on the sorter and the alphabetizing checked by hand.

3b. The collator is now used to insert a sequence number card in front of each set of cards for the same student. These sequence cards are prepared in advance and carry a control punch as well as numbers in sequential order from 0001 to 3999. Only odd numbers are used, the evens being reserved for future expansion due to errors or to late registration changes.

4b. The progress test cards with the inserted sequence cards are run through the tabulator with a comparing control on both sequence number and student number, with the machine set to tabulate. The resulting list is then checked visually to see that a sequence card has been inserted at the proper place and that the test cards are in proper alphabetical order. Any errors are corrected.

5b. The deck of cards used in number 4b is run through the reproducer, and the sequence numbers are punched from the sequence cards into the test cards.

6b. The progress test cards can always be placed in strict alphabetical order merely by sorting them on the numerical sequence number of four digits.

7b. Next, placement test deciles are gang punched into the test cards. This is done by sorting on student numbers (both the placement test decile cards and the tests cards carry student numbers), with decile card coming first, and running the entire deck through the reproducer with a control on a suitable punch in the decile card.

8b. When a progress test is given, the answer sheets are scored on the test scoring machine, the answer sheets distributed, and percentile ranks computed and recorded on the answer sheets. Letter grades are recorded also, if any are assigned.

9b. The answer sheets are alphabetized and checked against the deck of progress test cards for that course; cards are pulled for absentees; and both the answer sheets and the cards are put in the same order to facilitate punching the test results into the cards.

10b. Percentile ranks (and grades, if any) are punched from the answer sheets into the test cards on the alphabetical printing punch. The punching is checked.

11b. The cards now go to the tabulating department where the following operations are executed:

(1) Absentees are re-inserted on the collator and the sequence of the cards checked on the collator.

(2) An alphabetical list of all students is prepared on the tabulator showing the following data: student name and number, course and section, and the results of all progress tests to date.

(3) An alphabetical list of students by sections is run on the tabulator, showing the same data as for (2), above.

12b. The lists made in (2) of 11b are checked orally against the answer sheets as an added precaution to insure accuracy.

13b. The answer sheets and lists of results are given to the proper instructor for each section. The lists are for his use; the answer sheets are returned to the students. Also, the students receive key sheets of the correct answers, so that they may see just what they missed on the test.

14b. About once a month a composite list is run on the tabulator. This list shows the progress test results for all courses for each student, and through them it is possible to see how the student is doing in all of his courses. The sequence number is used to put the cards in one alphabetical order, where all the cards for each student are together. It has been found best to keep the cards in the order in which the composite lists will be run. When a test is given, the cards for that test only are selected from the entire deck. As soon as the details of handling that test are completed, the cards are collated back into the composite deck.

15b. By the end of the school year, the cards represent a complete picture of the record of each student in each course. All kinds of statistical studies are possible from these cards, among which are:

(1) Correlation between placement tests, progress tests, and comprehensive examination grades.

(2) Investigation of quality of work done by those who drop or resign.

(3) Correlation between grades in different courses.

(4) Grade distributions.

C. *Punch Cards and Comprehensive Examinations.*—At the end of the school year it is necessary to give, score, and report grades for all the comprehensive examinations within a period of less than two weeks. It has been found that the use of punch cards speeds the work and increases the accuracy. The steps in preparing and using these cards are:

1c. A deck of comprehensive examination cards is made for each course by reproducing the progress test cards, except that it is necessary to omit the first two progress test results to make room for the raw scores on the comprehensive. The remainder of the card is reproduced because the placement test results and the progress test results are useful as aids in setting the comprehensive grades.

2c. Decks of master cards for raw score intervals are prepared. These are used to enable the tabulator to make the distribution of raw scores for each examination. Decks with intervals of 2, 5, and 10 have been made. It has been found helpful to have duplicate decks to facilitate handling of courses where identical intervals are used. These cards carry a control punch.

3c. Percentile rank master cards are prepared also. An interval of 1 is used, and the range is from 01 to 99. It is well to have several sets of these and to have an abundance of cards for ranks below ten and above 90, since several class intervals may have the same percentile rank within these ranges. These cards carry a control punch and are used to gang punch percentile ranks into the examination cards.

Note: In all instances it is well to have the master cards with a corner-cut different from that of the detail cards.

4c. In handling comprehensive examinations the student number rather than the name is used to identify the student. This is done to impersonalize the examining and to simplify the procedures, because operations can be done more readily on a numerical than on an alphabetical basis.

5c. These cards are used to prepare attendance lists for each examination room and a master-list for use in checking in the papers at the end of the examination.

6c. As soon as the examination is over, the answer sheets are placed in numerical order and checked against the check-in rolls. This is done to make certain that no answer sheets have been misplaced. A special form is filled out for each absentee and inserted in its proper place in the stack of answer sheets. This means that there is either an answer sheet or an absentee sheet

for each name on the roll and for each examination card. This has been found to be preferable to pulling the cards for the absentee.

7c. The answer sheets are scored on the test-scoring machine. If more than the front of one answer sheet is used, all of the raw scores are recorded on the front of the first answer sheet. If the student has more than one answer sheet (which is usually the case, since most of the examinations have both morning and afternoon sessions), great care must be used to be certain that all the scores on the sheet are for the same individual.

8c. After the answer sheets have been scored and checked and the addition of the scores completed and checked, the total raw scores are punched into the cards mentioned in number 1c. The punching is facilitated because both the cards and answer sheets are in order by student number and there is either an answer sheet or an absentee blank for each card.

9c. From now on all operations are mechanical except the checking of the punching of the total scores. After the checking, the cards are sorted on total score, with the interval cards being placed first in the hopper. This sorting places the cards in order by score, with the interval cards coming at the proper place. It is well to check the sequence of the cards on the collator.

10c. The cards are now tabulated with a control being taken on the 11-zone punch in the interval cards (the detail cards being blank in the control column). The tabulator will print the intervals from the interval cards, count the detail cards between interval cards, and record the frequencies and the progressive totals. To insure that the count for each interval is placed on the same line with the intervals, only the upper hub of the comparing control must be used (i.e., there is no plugging to the add-hub of the control source), so that a control change will occur only when going from blank to X-punch, but there will be no change when going from X-punch to blank.

11c. The percentile ranks are computed from the progressive totals, and the proper percentile rank card (selected from the prepared deck of percentile rank cards) is inserted manually just back of the interval card with which that percentile rank goes.

12c. The cards are re-tabulated with both the interval cards and the percentile rank cards in the deck. In this operation the percentile rank card is handled as a detail card (i.e., it is blank in the control column), but it carries another X-punch to keep the tabulator from counting it. By taking the percentile rank through a counter, it is possible to print the ranks on the same line with the interval, the frequency, and the progressive total. Both interval cards and detail cards are blank in the percentile

rank column, so that all the counter receives is the percentile rank. This re-run serves as an accurate check on the original distribution and the insertion of the percentile ranks at the proper place. The first step of the check is to compare the second distribution with the first to check frequencies and progressive totals. The insertion of the percentile rank cards can be checked by comparing the machine-recorded ranks with those obtained in the computation on the first distribution.

13c. The complete deck of cards is run through the reproducer to gang punch the percentile ranks into the detail cards. It is not necessary to remove the interval cards before this operation if both the interval cards and the percentile rank cards contain a common X-punch which can be used to clear the punch magnets. It should be recalled that the percentile rank cards were inserted behind the interval cards. After the punching has been finished and sight-checked, both the interval cards and the percentile rank cards can be separated from the deck by using the sorter. Then the percentile ranks are interpreted.

14c. The deck of comprehensive examination cards is left in order by score and percentile rank until after the grades are set. Since the cards contain the placement test results, most of the progress test results, and the score and rank on the comprehensive examination, the information they reveal helps in setting the grades. Also, the distribution of raw scores and the specified answers given to certain key questions are used in setting the grades. For example, in setting the grades someone may wonder what kind of persons we find at the 10th percentile from the bottom. By referring to the comprehensive examination cards, we can learn the quality of their placement test results and their relative success on progress tests, and by referring to the answer sheets we can see which questions they missed and which they answered correctly. If we find that most of the students at this percentile are missing items based upon elementary facts and principles, we feel that we cannot pass persons at that level. A higher level can be investigated until a satisfactory one is found. Such a procedure can be used until all of the division points for the various grades have been set.

15c. After the grades have been set, the grades are gang punched into the detail cards and the grades interpreted.

16c. Next the detail cards are alphabetized on the sorter by sorting on sequence number, and grade reports are printed by the tabulator on specially prepared forms, so that copies of the grades can be reported by the Registrar and to others concerned.

17c. Since the grade cards contain so much pertinent information regarding each student's academic record during the year, many statistical studies can be made from them.

D. *Punch Cards and the Library of Test Items.*—An item analysis is made of the items used on each progress test and comprehensive examination. In making the analysis, 100 answer sheets uniformly distributed throughout the highest and lowest quarters are selected as a sample. (If the number of examinees is less than 400, samples of 50 may be used, or the upper and lower halves may be used instead of quarters.) The analysis is made on the test-scoring machine by utilizing the graphic item counter, and the count is made on the correct responses. For the items found to have low discriminating value the frequency counts for each distractor are made, unless the low validity is obviously due to excessive ease or difficulty. Four measures are secured for each item: V, the validity or discriminating power, which is the tetrachoric coefficient of correlation between the item and the total test; D, the difficulty, expressed as per cent of the group answering the item correctly; H, the per cent of the highest quarter or upper half answering the item correctly; and L, the per cent of the lowest quarter or lower half answering the item correctly.

An outline is made of the course content of each comprehensive course with major, intermediate, and minor classification of topics, and on the basis of this outline each test item is classified according to the aspect of the course which it covers. This classification and the item analysis data are all punched into cards, and the statement of each item is typed on the back of the card. This provides a very complete and usable library of test items. The actual steps in preparing the cards for this library are:

1d. Pre-punching a deck of cards to show the course and the date the test was given.

2d. Punching the validity data (V, D, H, and L) and the course content into the cards prepared in 1d. One digit is used for each validity datum, i.e., a validity correlation coefficient of .30-.39 is written as 3 and for the per cents for D, H, and L

the units value is dropped; for example, a per cent of 60-69 is punched as 6. This is done to conserve columns on the cards. (Of course, this procedure will mean that the punched values will average about .05 or 5 lower than the computed values.)

3d. The test item involved is typed on the back of the card which carries the corresponding item analysis data.

4d. The typing is proofed, the correct answer indicated, and the item is filed according to test content.

The information which is punched into these cards makes it possible to select mechanically items according to validity, difficulty, and course content and to make various types of statistical studies.

MACHINES IN CIVIL SERVICE TESTING

SIDNEY W. KORAN

Employment Board, Pennsylvania Department of Public Assistance

EDITOR'S NOTE: This article is an abridged version of a lecture recently presented by the author before an in-service training class comprising staff members of the examination division of a large state civil service agency. It is offered here because it pulls together, for what is probably the first time, numerous loose ends of the important body of knowledge that is beginning to come into existence on the mechanization of civil service examining processes. The article comprises a description of the purpose, design and operation of the I.B.M. scoring machine, a discussion of the limitations of the scoring machine in connection with the conduct of examinations, information on adapting tests to machine scoring, descriptions of procedures for scoring tests using the I.B.M. and other machines, information on scoring various types of rating scales by machine, material on the uses of the scoring machine in item analysis and in the computation of several statistical measures, and a summary of the place of tabulating equipment in the conduct of certain examination tasks.

The I.B.M. Scoring Machine

The I.B.M. scoring machine was designed to meet a very real and pressing need in the field of educational and psychological testing for a method of scoring objective tests that would combine speed, accuracy, and low cost to an extent considerably beyond that possible with any technique previously developed. The recognition of that need was sufficiently great to stimulate, over a period of years, the development of numerous ingenious techniques as well as several different types of automatic and semi-automatic devices. The one, however, which appears to have been the most successful and to have earned the widest application to civil service use is the comparatively new I.B.M. scoring machine.

The operation of this machine is dependent upon the principle that the mark made by a pencil having a soft lead will conduct electricity. In order to score a test by the machine it must be designed so that the examinee may indicate his answers to the questions by placing pencil marks in certain predetermined and properly labeled positions on a sheet of paper which is either separate from the test booklet or may later be separated from it. For consistently satisfactory results it is advisable to furnish examinees with mechanical pencils equipped with special high-graphite-content leads and to use answer sheets that have been carefully and accurately printed so that the location of each one of the 750 possible response positions on either side of the sheet corresponds within fairly close limits to the location of each of the 750 sets of contacts within the machine.

Each of these sets of contacts consists of five small parallel blades insulated from one another and connected alternately to the positive and negative sides of the electrical circuit, the current for which is furnished by several conventional radio "B" batteries. When an answer sheet is inserted in the machine for scoring, it is pressed against a plate containing the 750 sets of contacts. Whenever one of these sets of contacts presses against a pencil mark, the latter, since it is a conductor, permits current to flow across one or more of the four gaps made by the five parallel blades. The length of the pencil mark determines whether one, two, three, or four of these gaps will be bridged. If the examinee follows instructions and makes his pencil mark sufficiently long, it will bridge all four of the gaps. By designing the machine so that there are several millions of ohms of resistance in series with each set of contacts, current differences which result when some of the examinee's marks are not long enough to bridge all four gaps are minimized sufficiently to prevent their having an appreciable effect on the score.

The resistance in each circuit is such that when the appropriate rheostats have been adjusted properly, a single unit of

current is registered for each set of contacts that may be pressed against a pencil mark. Scores are read on a meter which has been calibrated in terms of these units. The use of switches and other accessories makes it possible to read rights, wrongs, omits, rights minus wrongs, rights plus wrongs, rights minus or plus a fraction of wrongs, etc. Whether any given choice, or answer position, will be counted as right or as wrong, or whether it will be eliminated from the scoring altogether, is determined by the manner in which holes have been punched in the set of keys inserted in the scoring rack. By the use of switches and the proper perforation of field selection holes in the scoring key, the machine may be adjusted so that the meter will read the score for all of the items on the answer sheet or for certain combinations of the ten 15-item fields, or both.

Machine-Scorable Answer Sheets

Standard answer sheets designed to fit several types of general situations are available from the manufacturer of the machine, and it is also possible to have special answer sheets, designed to meet certain specific requirements, printed to order. Unless ordered in fairly large quantities, however, special answer sheets are usually more expensive than standard answer sheets and their use frequently introduces an additional time element into the planning of examinations.

Some of the agencies, in an effort to take advantage of the economies afforded by quantity purchases and to enjoy the convenience of having their own stock on hand, have standardized their major requirements sufficiently to permit them to order relatively large quantities of three or four types of answer sheets printed only with the name of the agency, the item numbers, and the response positions. As new examinations come up, these agencies select the type of answer sheet which most nearly fits the requirements of the particular situation and print or multilith, in the left-hand margin, whatever additional identifying material is required.

Limitations Imposed by Machine Scoring

To the individual who is about to construct a test that is to be scored by machine the limitations imposed by the machine method are chiefly three: 1. A separate answer sheet must ordinarily be used. 2. The response to each question must be indicated by making a special kind of pencil mark. 3. The orientation of the response positions on the answer sheet cannot be altered.

The first of these considerations, that of the use of a separate answer sheet, is more correctly a "condition" rather than a "limitation," for even when the machine method of scoring is not involved it is usually desirable, when examinations comprising large numbers of items are to be administered to any considerable number of individuals, to make use of some form of separate answer sheet in order to facilitate the scoring process. This is true whether it is planned to do the scoring entirely by hand or by some combination of hand-scoring and overprinting (with a multilith, for example).

Another reason why the use of a special answer sheet is not ordinarily a serious obstacle is that it is frequently possible, if necessary, to design the examination so that the test questions are printed directly on the answer sheet beside (or directly over or under) the response positions. This has been done in the case of a number of standardized educational and psychological tests and has also found some, though much more limited, use in connection with civil service examinations. While this procedure seems to be particularly advantageous when used with one- or two-page personality inventories and, as will later be pointed out, certain types of rating scales, there are ordinarily several objections to its routine use in setting up civil service examinations. Among the principal objections are the increased trouble and expense caused both by the special printing requirements and by the fact that the relatively small number of items which may be printed on a letter-size sheet usually necessitates using several answer sheets for a test of any appreciable length. The handling, scoring, and computational difficulties encountered whenever a single test requires more

than both sides of one answer sheet are ordinarily sufficient to discourage that practice. There are, however, certain situations in which the mere fact that several answer sheets will be required for a given test may well be a matter of relatively minor importance when considered beside the larger aims of the examination.

The second limitation imposed by the use of the scoring machine method, that of the necessity for making a special kind of pencil mark to indicate the answer to each question, is apparently proving less of a problem than many expected it would. As returns come in on the results of research (1) and on the development of improved scoring procedures which have reduced the likelihood of scoring errors to a negligible figure (3, 11), it is clear that whatever problem is actually presented by this limitation may be pretty adequately neutralized by taking the following three steps:

1. Include in the examination announcement a section comprising (a) an explanation of the types of questions that will be used, (b) a statement of the fact that the test will be scored by a machine which will provide the correct score if the examinee follows all instructions carefully, (c) a list of the rules which must be followed in indicating the answers, and (d) a group of sample questions printed alongside a specimen portion of the answer sheet on which the answers to some of the sample questions have been properly marked and the remainder left for the examinee to complete as a practice exercise.[1]

2. Provide the examinee, at the time the examination is administered, with a page of instructions, similar to those described above, which he is given time to review sufficiently long to refresh his memory and to which he may refer at any time during the examination. (One example of such an approach is the *Directions for Using the Answer Sheet,* reproduced on the following pages, which has been used in Pennsylvania since 1939 by the Employment Board of the Department of Public Assistance with examinations conducted for approximately 115,000 persons.)

3. Furnish, throughout the test, additional adequate and clear directions including, wherever a somewhat different approach is employed, a sample question properly answered (7). (Several illustrations of such special directions appear among the examples of test material presented later in this article.)

[1] An example of this approach is the page entitled *Sample Questions for the General Test* which appears as part of current U. S. Civil Service Commission announcements for examinations which will include machine-scored written tests.

DIRECTIONS FOR USING THE ANSWER SHEET

All of the answers in the test you are about to take are to be recorded on special ANSWER SHEETS instead of in the Question Booklet. *To receive credit for your answers they must be recorded in the proper spaces on your* ANSWER SHEET.

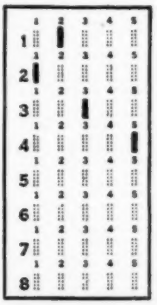
ANSWER SHEETS will be scored by an electrical test-scoring machine. In order for your test to be scored accurately, it is necessary for you to observe the following directions carefully:

1. Read each question and its numbered answers and decide which answer is correct.
2. Find the pair of dotted lines numbered the same as the answer you have chosen as being correct, and blacken this space with your pencil. *Be sure that the space you mark is in the row numbered the same as the question you are answering.* Misplaced answers are counted as wrong answers.
3. Indicate each of your answers with a vertical solid black pencil mark. Solid black marks are made by going over each mark two or three times and by pressing firmly on your pencil.
4. If you change your mind, *erase your first mark completely, then mark the correct space.* Blacken one space *only* for each question number.
5. Do not rest the point of your pencil on the ANSWER SHEET while you are considering your answer and *do not make unnecessary marks.*
6. Keep your ANSWER SHEET on a hard surface while marking your answers.
7. Make your marks as long as the pair of dotted lines.

Below are some sample questions to give you practice in using the ANSWER SHEET. The questions at the left are similar to those you will find in your Question Booklet. At the right is an illustration of a portion of an ANSWER SHEET. The answers to the first four questions have already been marked on the ANSWER SHEET. Study the questions and note the way the answers to them have been marked on the ANSWER SHEET. Then answer each of the remaining questions in exactly the same way; that is, by making a heavy black mark on the ANSWER SHEET in the space numbered the same as the correct answer.

Questions for Practice

1. The third month of the year is: (1) February; (2) March; (3) January.
2. The capital of Pennsylvania is: (1) Harrisburg; (2) Albany; (3) Boston.
3. The Governor of Pennsylvania is: (1) John Garner; (2) Alfred Landon; (3) Arthur James.
4. If one pencil costs one cent, five pencils will cost: (1) three cents; (2) six cents; (3) four cents; (4) two cents; (5) five cents.
5. George Washington was the first President of the United States. (1) True; (2) False.



6. Every calendar year has: (1) eleven months; (2) ten months; (3) twelve months.
7. The fuel most commonly used in automobiles is: (1) kerosene; (2) carbona; (3) crude oil; (4) gasoline.
8. The sum of six and four is: (1) five; (2) six; (3) eight; (4) nine; (5) ten.

NOTE: The Answer Sheet provides spaces for recording five different choices for each question. Some of the questions in your examination booklet may contain only two, or three, or four choices. When answering a question containing fewer than five choices, you are to ignore the additional spaces printed on the Answer Sheet for that question.

The last of the three limitations mentioned as imposed by the machine method of scoring has to do with the fact that the relative location of each of the 750 response positions on the answer sheet is fixed and may not be changed by the test constructor. It is in meeting this difficulty that the test technician's ingenuity enters the picture.

Despite the handicaps of perpetually imminent deadlines and understaffed examination units—two well-known characteristics of the conditions under which many civil service commissions work—a sufficient number and variety of adaptations of test material to this particular limitation of the machine method have been produced in the short space of a few years to warrant the conclusion that the use of separate answer sheets involving fixed response positions offers no particularly serious obstacle to the construction of objective test material. In fact, what used to be a double bottleneck—construction and scoring—has, because of the advantages offered by the scoring machine, been reduced to but a single obstruction. Objective tests may now be scored so cheaply that the major remaining obstacle to the wider use of *better* objective tests appears to be the difficulty of constructing them under the usually prevailing conditions of insufficient time and the not-too-ready availability of trained examination technicians.

Constructing Items for Machine Scoring

In general, the basic rules to be followed and the pitfalls to be avoided in the construction of good objective-type items

are as applicable to items that are to be scored by machine as they are to items that are to be answered either directly in the test booklet or on a separate sheet designed to be scored manually. An item that is "tricky" or that contains an ambiguous or ludicrous statement is unacceptable for reasons quite apart from the scoring method that will be employed. While there are certain additional points that must be kept in mind—mostly with regard to adequate instructions to the examinee and strict adherence both to the physical limitations of the answer sheet and the electrical limitations of the machine itself—fundamentally, an item that is unsatisfactory for any reason that would interefere with its suitability as an ordinary objective-type question is equally unsatisfactory for use in a machine-scorable test. Here, however, are some of the considerations which appear to be sufficiently peculiar to the use of machine-scorable answer sheets to warrant enumeration and brief discussion:

1. *Choice of answer sheets.* Wherever practicable, test items should be designed to make use of standard answer sheet forms. Doing so keeps down construction time as well as costs and obviates the necessity for presenting special instructions not covered by the general directions printed in the announcement and furnished the examinee at the time of the examination. The use of standard answer sheets possesses the additional advantage of capitalizing on the fact that, since machine-scored tests are being used more and more widely by civil service agencies and educational institutions, an increasingly large proportion of the civil service test-taking population may be expected to have sufficient previous experience with standard forms of separate answer sheets to permit them to concentrate on the test material with a minimum of distraction and tension.

2. *Adequate instructions.* Instructions for indicating answers to such specialized subtests as those involving alphabetizing, proof-reading, checking, sorting, filing, punctuation, and similar tasks should be adequate and should preferably include a sample exercise properly answered. In writing these

directions the kind of language, specificity, and need for examples will, of course, vary according to the level of the job for which the examination is being designed. In general, however, it is safer to be somewhat too specific and to provide what, to the sophisticated test-taker or Ph.D. test constructor, may sometimes appear to be an unnecessary example, than to take too much for granted on the part of the examinee.

It is sometimes argued that an examinee who can't follow such simple instructions "doesn't deserve to pass" or "couldn't do the job anyway." While there are certainly times when this stand appears justifiable, the writer's opinion is that it is always safer to provide, if for no other reason than the maintenance of satisfactory public relations, directions that meet the highest standards of adequacy. If it is desired to test the examinee's ability to follow instructions, one should use a test designed to do just that, rather than take the chance of measuring such a trait by using "complicated" (in this case, a euphemism for "inadequate" or "unsatisfactory") instructions which are likely to result in a distribution of scores unduly influenced either by the degree of an examinee's previous experience with new types of tests or by the extent to which he exhibits caution in situations of this kind.

In this connection it sometimes occurs, in construction of a test for a position such as building superintendent, that by the time the test constructor has finished adapting a certain bit of practical material to the limitations of the answer sheet, his product, regardless of how cleverly worked out, is no longer suitable for that particular level of job. What he has may be an excellent test of intelligence, but of a higher level than that required of a building superintendent. It hurts, sometimes, to have to discard or extensively modify a brain child of that kind, but it has to be done.

3. *Item sequence.* The sequence of items in the test should be such that subtests or item-groups that are to be weighted differently from other portions of the test or for which a separate score will be desired, will fall entirely within one or more fields. If, for example, in a test for key punch operators

a separate score will be required for a 30-item subtest on the subject of coding, the items for the entire test of which the coding items are a part should be arranged so that the item numbers (if a standard answer sheet is used) will start with 1, or 16, or 31, or 46, etc. When the test is being scored it will then be possible, if the proper field selection holes have been punched in the answer key, to read the score of the subtest with the expenditure of no more additional effort than is required for turning a knob while the answer sheet is in the machine. Similar treatment should be accorded item groups to which a correction formula is to be applied that is different from that used for any other part or parts of the entire test.

4. *Reducing response errors.* Care should be taken to avoid wording questions and selecting styles of type or print that are likely to cause the examinee to make unnecessary clerical errors in recording his responses. This not infrequently occurs, for example, when Arabic numerals having the same range as those used to denote response positions are used in the answer. Answer sheets are available which eliminate this difficulty by using the letters A, B, C, D, and E to designate the response positions. Where the response positions are numbered, however, it is frequently helpful simply to spell out the numbers from 1 to 5 when they appear alone or almost alone in the answer. Two simple illustrations of this point are:

> "The number of inches in one-third of a foot is: (1) two; (2) three; (3) four; (4) five; (5) six." *instead of:* "(1) 2; (2) 3; (3) 4; (4) 5; (5) 6."
>
> "How many persons in the family are eligible to receive some form of assistance? (1) none; (2) one; (3) two; (4) three; (5) four." *instead of:* "(1) 0; (2) 1; (3) 2; (4) 3; (5) 4."

5. *Juxtaposition of instructions and related items.* Where use is made of a page of instructions including a key, legend, code, or similar device likely to have to be referred to frequently by the examinee in order to answer a given group of related questions, the format of the examination booklet should be such that the page containing the instructions will

face at least a group of the questions. Among the possible exceptions to this rule is the situation in which one of the functions being tested is the ability to memorize certain material or relationships, and in which the test is being timed in an effort to obtain a measure of the examinee's ability to do so.

6. *Completion arithmetic items.* The construction of choices for items involving arithmetic, algebraic, or statistical problems, or consulting a graph or chart, is no different when the item is to be scored by machine than by any other method. There may, however, be situations in which it is considered important, in connection with a certain group of items in a test, to know the *exact* answer arrived at by the examinee as a result of his calculations. When this is so, it is possible by expending some additional time and effort, to retain the advantages of the completion type test for this particular group of questions and at the same time have the machine-indicated score represent the examinee's achievement in the entire test. This may be accomplished by designing the answer sheet so that spaces are provided both for the examinee to write in his answers to the questions and for a clerk to indicate the correctness of those answers by making the usual kind of pencil marks in response positions especially provided for that purpose.

Two disadvantages of this approach are the need for special answer sheets and the time and expense involved in having the completion items scored manually before the test as a whole can be scored by machine . Another possible disadvantage is the effect that the use of two types of directions may have on the examinee's adherence to the important and oft-repeated general instructions to "indicate your answer to each question by making a heavy mark in the appropriate space on the answer sheet." This is of some importance, for the extent to which the examinee can be persuaded to accept the idea of making proper marks instead of writing answers or numbers on the separate sheet is one measure of the amount of machine vs. handscoring that will have to be done. For these reasons, and because it is probably possible to use the multiple-

choice form of presentation for arithmetic and similar items without interfering seriously with their validity, it would seem preferable, ordinarily, to avoid mixing the two types of responses in the same test.

Beginning on the next page are examples of Test Material Adapted to Machine Scoring.[2]

Scoring Civil Service Tests with the I.B.M. Machine: Historical Note

Civil service commissions, while recognizing the speed and economy features of the machine-scoring method of rating objective tests, were at first quite reluctant to put the machine to use. Springing in part, probably, from the usual resistance to adopting methods and procedures differing from those already in use, the criticism was made that not only did the machine method involve the use of special materials to which the test-taking public might object, but the scores which it turned out were insufficiently accurate.

This controversy was quietly running its course when an event occurred in the field of public personnel administration that resulted, among many other things, in removing the whole question from the *talking* to the *trying* stage. Suddenly, all over the country, sizeable civil service agencies were coming into existence in accordance with the merit system provisions of the Social Security legislation. Many of these new commissions were faced with the task of examining unprecedented numbers of persons within time and budgetary limitations that were not easy to meet, and some were composed of commissioners and administrators who were sufficiently new to civil service problems to be relatively quite receptive to such new-

[2] These examples of machine-scorable subtests and item-groups are offered solely for the purpose of illustrating the variety of test material that may be adapted to machine scoring and the kinds of instructions that may be employed. The writer wishes to thank the Employment Board of the Pennsylvania Department of Public Assistance and Miss Hilda P. Thompson, Executive Director, for their kind permission to use this material, which was developed for the Board over a period of several years by C. R. Adams, G. K. Bennett, S. W. Koran, B. V. Moore, E. A. Rundquist, and K. S. Wagoner. C. H. Smeltzer and M. S. Viteles were employed as consultants to the Board during the period this material was being developed.

CIVIL SERVICE TESTING

NUMBER AND NAME CHECKING

In each of the following groups, some of the pairs of names and numbers are exactly the same while others are different.

You are to check, on the connecting line, the pairs which are different and indicate on your ANSWER SHEET the total number of such pairs in each group.

EXAMPLE:

61.	John Smith	John Smith
	Wm. G. Burns	Wm. C. Burns
	Thos. Doe and Co.	Thos. Doe Co.
	Burt Salt Corp.	Burt Salt Corp.
	Bryant, Mitchell	Bryant, Mitchel

In the example above, three pairs are different so you are to make a heavy mark in space number 3 opposite question No. 61 on your ANSWER SHEET, thus:

61 1 2 **3** 4 5

Be sure you have marked No. 61 on your ANSWER SHEET, then go on to No. 62. Work as fast as you can without making mistakes.

62.	Auto Service Shop	Auto Service Shop
	Alester & McAlester	Alester & McAlester
	Van & Van De Vyere	Van & Van de Vyere
	Raymond Pretschold	Raymond Pretschold
	Kassanitsch Storage	Kassanitsch Storage

62

63. Paul A. Anderson — Paul A. Anderson
Knight & Orgto — Knight &

83.	17287	17287
	27355453231	27355453731
	829	829
	53728	53728
	3212	3212

83

84. 736291 — 736291

ALPHABETIZING

Rearrange the names in each of the following groups in the order in which they would appear in an alphabetic file.

List the names in alphabetic order in the blanks at the right. The number in parentheses after each name must be kept with that name during the alphabetizing. When you have arranged the names in the blanks, indicate the alphabetized order by making a heavy mark on your ANSWER SHEET in the appropriately numbered space opposite the proper question number.

For example, when you alphabetize the names in the first group below, you will find that the name followed by the number 3 in parentheses belongs after No. 46. You will therefore make a heavy mark in space number 3 after question No. 46 on your ANSWER SHEET, thus:

46 1 2 **3** 4 5

Alfred Anthony	(1)	46.	Anthony, A.L. (3)
Estelle Anthony	(2)	47.	
A. L. Anthony	(3)	48.	
Emily M. Anthony	(4)	49.	
Miss Birdie Anthony	(5)	50.	

NUMBER, NAME, AND ARITHMETIC CHECKING

DO NOT OPEN THIS BOOKLET UNTIL GIVEN THE SIGNAL BY THE PROCTOR

Read the following directions very carefully:

The inside pages of this booklet contain three tests which will be timed. The first test consists of pairs of numbers and the second test of pairs of names. If the numbers or the names of a pair are exactly the same, make a heavy mark in space number 1 on your ANSWER SHEET beside the corresponding number of that pair. If they are different, make a heavy mark in space number 2 on your ANSWER SHEET.

The third test consists of simple arithmetic calculations. If a problem is correct, make a heavy mark in space number 1 on your ANSWER SHEET beside the corresponding number of that problem. If it is not correct, make a heavy mark in space number 2.

SAMPLES DONE CORRECTLY

Question Booklet	Answer Sheet
1. 453----------435	1
2. 6125----------6125	2
3. William Johnson----------William Johnson	3
4. Abraham and Link Co.-----Abraham and Lind Co.	4
5. 4 + 8 = 12	5
6. 3 x 6 = 15	6

NOW DO THE SAMPLES BELOW

7. 326----------326	7
8. 7418----------7814	8
9. Samuel Dillon----------Samual Dillon	9
10. Markwell and Gordon-----Markwell and Gordon	10
11. 14 x 5 = 70	11
12. 12 ÷ 6 = 3	12

Whenever the proctor says "Stop," STOP WORK and listen carefully for further instructions.

FAILURE TO FOLLOW INSTRUCTIONS EXACTLY MAY LOWER YOUR SCORE IN THE EXAMINATION.

DO NOT OPEN THIS BOOKLET UNTIL GIVEN THE SIGNAL BY THE PROCTOR.

Make a mark in space number 1 if the numbers are exactly the same.
Make a mark in space number 2 if the numbers are different.

1. 616----------626
2. 4572----------4752
3. 0618----------0618

Make a mark in space number 1 if the names are exactly the same.
Make a mark in space number 2 if the names are different.

1. John L. Frankson----------John L. Franksan
2. Overholt Tobacco Co.----------Overholt Tobacco Co.
3. Time Recording, Inc.----------Time Recording, Inc.

Make a mark in space number 1 if the calculation is correct.
Make a mark in space number 2 if the calculation is incorrect.

1. 13 + 5 = 18
2. 9 + 6 = 15

CIVIL SERVICE TESTING

ALPHABETIZING AND SORTING

Each name below represents an addressed letter. You are to determine the number of letters addressed to each person. First, write the names of the Junior Visitors, Senior Visitors and the Interviewers in alphabetical order in the spaces provided. Then tabulate on the spaces provided the letters each received. Make a heavy mark on the ANSWER SHEET to indicate this number. For example, if the person received 3 letters make a heavy mark in space 3 beside the number of that name on the ANSWER SHEET. Thus, since Juanita Bates is the name which will be first when the Junior Visitors are arranged in alphabetical order, this name is written at the top of the Junior Visitor list. Tallying will show that she received two letters. Hence you will make a heavy mark in space number 2 beside Question No. 61 on the ANSWER SHEET.

Name	Position
Alberta Cummins	Junior Visitor
Alfreda Swift	Junior Visitor
Helen Cushman	Senior Visitor
Rita Bauman	Junior Visitor
Alberta Swift	Senior Visitor
Mary Petrey	Senior Visitor
Eleanore Petry	Junior Visitor
Rose Bowman	Senior Visitor
Alberta Swift	Senior Visitor
Juanita Bates	Junior Visitor
Jenny Betts	Senior Visitor
Jeanne Bolten	Interviewer
Alberta Cummins	Junior Visitor
Mary Beck	Interviewer
Mary Petrey	Senior Visitor
Kathryn Snow	Interviewer
[illegible] Bauman	Junior Visitor
[illegible]	[illegible]

	Junior Visitor	Space for Tallying
61.	_____	____
62.	_____	____
63.	_____	____
64.	_____	____
65.	_____	____

Senior Visitor

SORTING

This is a sorting test. The city in which each person lives is represented by a code symbol. Determine the number of persons living in each city.

Use the blanks provided in the code list for purposes of counting. If the code symbol for a city is not listed, count it as Miscellaneous (No. 30). When you have finished sorting the names, count the number living in each city and indicate the total for each city by making a heavy mark in the proper space on your ANSWER SHEET.

For example, if it is found that four persons live in a city whose code symbol is HD-4, you would make a mark in space number 4 beside the question number of that code symbol, thus: 1 2 3 4

APPLICANT'S NAME	CODE FOR CITY	APPLICANT'S NAME	CODE FOR CITY
Moore	RB-8	Gray	JR-3
Evans	SS-4	Cooper	FS-1
Foster	MB-6	Force	HB-7
Brown	HB-7	Burtt	PT-2
Jones	PT-2	Lewis	RB-8
[illegible]	AK-4	[illegible]	[illegible]
[illegible]	[illegible]	Arno	[illegible]
Crown	[illegible]	Murphy	WY-5
Wells	MB-6	Kahn	VB-8
Fink	KL-9	Rolfe	QR-1
Sporn	QR-1	Borg	KL-9
Call	WY-5	Hansen	FS-1
Swift	SS-4	Rees	RB-8
Harris	MB-6	Roth	UM-7
Moses	ER-5	Prince	MB-6
Brand	FS-1	Creel	SS-4
Moon	AK-4	Yule	PT-2

CODE LIST

16. AB-3_____
17. AK-4_____
18. ER-5_____
[illegible]
24. PT-2_____
25. QR-1_____
26. RB-8_____
27. SS-4_____
28. VB-8_____
29. WY-5_____
30. Miscellaneous____

SPELLING

The following letter contains a number of misspelled words. Cross out each misspelled word. Indicate on your ANSWER SHEET the number of misspelled words in each line by making a mark in space number 1 if one word in the line is misspelled; in space number 2 if two words are misspelled, and in space number 3 if none of the words in the line is misspelled.

61. It is a pleasure, unparaleled in my experience to recommend 61

62. John Gains, who has been my factory superintendant for ten years. 62

63. His imperturbible good nature and incredable energy have made him 63

64. my most valued associate. Independent in his views, he is 64

65. [illegible] accomodating to the needs of [illegible]cation. As [illegible] 65

CHECKING FOR ACCURACY

Below is a list of the names and addresses on a temporary pay-roll, grouped according to the city in which each employee lives and arranged in order of the amount owed. Some of these names appear on the following page arranged in alphabetic order. You are to compare each name, street address, city and amount of salary check with the same information listed on the next page. Pay no attention to names which are omitted from the table.

On the next page, check the names, street addresses, cities, and salaries which are not exactly the same as those on this page. Then make a heavy mark in the proper space on the ANSWER SHEET to indicate the total number of errors in each employee's record, thus: (1) one; (2) two; (3) three; (4) four; (5) none. Since but one error is counted for each name, street address, city, or salary in which errors occur, there can be, at the maximum, four errors in a record.

Name	Albany	Salary	Name	Geneseo	Salary	Name	Canastota	Salary
Peter Meteer		$99.90	Thomas B. Jarboe		$227.40	Herman Cutler		$300.00
3026 Belmont Place			1316 East Mentor St.			2713 Dimmith Ave.		
W. P. Ickes		76.90	John Frederick Rider		91.04	A. D. Waphington		98.05
1219 Perry Ave. N.			105 4th St. S.E.			2824 Garfield Ave. S.		
Alfred Robertson		58.85	E. Edward Edwards		84.50	Petroff Biesten		22.00
711 4th Ave. S.E.			1672 Kennedy Road			1237 Goddard Road		
			Pierot Maupassant		60.25	Montaigne Mattleton		8.45
	Binghamton	462.70	1838 20th Ave. N.E.			1204 Laidlow		

EXAMPLE: In question number 151 you will find that there is an error in the name (the first initial should be W), an error in the street address (the house number should be 3119), and an error in the city (it should be Delhi). A check has been made in each of the proper columns at the right of the table to indicate these errors. You will, therefore, make a heavy mark in space number 3 after question number 151 on your ANSWER SHEET to show that there are 3 errors in the record on the first line of the list.

	EMPLOYEE	STREET ADDRESS	CITY	SALARY	Name	Street	City	Salary	
151.	Anker, L. Paul	6119 Hartford St. N.W.	Binghamton	$347.95		✓			151
152.	Bachelor, Rosemary	413 East Ninth St.	Canastota	21.45			✓		152
153.	Biesten, Petroff	1237 Goddard Road	Canastota	22.00					153
154.	Clasp, Arcosoric	1901 Dale St. S.W.	Ellenville	50.00					154
155.	[illegible] J. S.	2753 [illegible]wood Ave.	[illegible]ghamton	27.65					155

PROOF READING

Reproduced below is a correct copy of a page of printing. On the right-hand page is a copy of the same material containing all kinds of errors. Count the number of errors in each line of the copy on the right-hand page. Make a heavy mark on the ANSWER SHEET in the space having the same number as there are errors in the line, thus: (1) one error; (2) two errors; (3) three errors; (4) four errors; (5) no errors. Every deviation from the copy below (except length of line) is to be counted as an error. Only one error is to be counted for each word or number group in which errors occur. If you check each error as you come to it, you will find it easier to count accurately the number in each line.

The total number of applications for public assistance received during the last quarter of 1938 was 194,743. This represents an increase of 6.1 percent over the total for the previous quarter and a rise of 3.9 percent as compared with October-December 1937.

Home relief applications received, numbering 177,102, were 4.9 percent above the figure for the preceding quarter and 3.6 percent

106. The total number of aplications for publi c assistance received 106
107. during the last quarter of 1938 was 194743 This represents 107
108. an increase of 61 percemt over the total for the previous quarter 108
109. and a rise of 3.7 per-cent as compared with November-December 1937. 109
110. Home relief applications received, numbering 177,102, were 110
111. 4.9 percemt above the figur for the preceeding quarter and 3.6 111
112. percent above the number received in the last three months of 112
113. 1937. nearly 9.1 percent of all applications filled were acount- 113

PARAGRAPH MEANING

DIRECTIONS FOR ANSWERING QUESTIONS 121 to 146 INCLUSIVE: Each paragraph below includes one word which spoils the meaning of the paragraph. This incorrect word is one of the five words which have numbers printed just above them. When you have found the incorrect word, make a heavy mark on your ANSWER SHEET in the space having the same number as the incorrect word.

1 2
121. Our present defeat by the machinery around us is a permanent
3 4 5
thing, a plateau in our progress to a slaveless world. 121

1 2
122. The development of the higher organisms may be regarded as due
3 4
to the sewing together of one of these cells to form a

FOLLOWING DIRECTIONS

This is a test of your ability to follow directions.

You are to classify the employees of a department according to the salary each receives. The schedule of classifications is as follows:

Class 1: $1000 to 1099;
Class 2: 1100 to 1199;
Class 3: 1200 to 1299;
Class 4: 1300 to 1399;
Class 5: below $1000 or above $1400.

The salary for an account clerk is $1100.
The salary for a mail clerk is $1025.
The salary for a stenographer is $1175.
The salary for a secretary is $1350.

A Junior employee in any of these positions receives $50 less than the amount shown, while any Senior employee receives $100 more. For example, a Senior Stenographer receives $1275 ($100 more than a Stenographer) while a Junior Stenographer receives $1125 ($50 less than a Stenographer).

After reading the directions you are to classify the positions listed below according to the salary each receives, with these exceptions:

1. Positions not mentioned in the above directions are to be placed in class number 5.
2. Individuals having five or more years' experience are to be placed in class number 5.
3. Individuals with less than 2 years' experience are to be placed in class number 5.

Indicate your answer by making a heavy mark on the ANSWER SHEET in the space having the same number as the salary classification.

EXAMPLES

A Stenographer with 3 years of experience will fall into Class 2, so a heavy mark is made in space number 2 on your ANSWER SHEET, thus:

A Senior Stenographer with 7 years of experience will fall into Class 5, so a heavy mark is made in space number 5 on your ANSWER SHEET, thus:

Question Number	Departmental Division	Position	Years Experience	Question Number
91.	Accounting	Account Clerk	4	91
92.	Pay-roll	Stenographer	3	92
93.	Administrative	Senior Stenographer	5	93
94.	Filing	Junior Secretary	2	94
95.	Clerical	Mail Clerk	7	95
96.	Filing	Junior Account Clerk	5	96
97.	Mailing	Stenographer	6	97
98.	Clerical	[illegible]		98

T-F PAIRS IN FIVE-CHOICE FORM

This part of the examination consists of 50 questions, each made up of two statements, labelled A and B. You are to determine the truth or falsity of each of the statements. Having done so, you are to indicate your answers on the ANSWER SHEET as follows:

1. If you consider that the answer to either or both statements in any question cannot be known, make a heavy mark in space number 1 on your ANSWER SHEET.

2. If you consider both statements in any question to be true, make a heavy mark in space number 2 on your ANSWER SHEET.

3. If you consider the first statement to be true and the second statement to be false, make a heavy mark in space number 3 on your ANSWER SHEET.

4. If you consider both statements to be false, make a heavy mark in space number 4 on your ANSWER SHEET.

5. If you consider the first statement to be false, and the second statement to be true, make a heavy mark in space number 5 on your ANSWER SHEET.

For your convenience, these directions are summarized below:

Mark 1, if none of the answers below applies.

Mark 2, if both statements are true.

Mark 3, if first statement is true, second is false.

Mark 4, if both statements are false.

Mark 5, if first statement is false, second is true.

EXAMPLE

(A) All public agency employees are happy.
(B) Federal grants to States for public assistance will be drastically changed by 1950.

In the above question, the first statement is false. However, nobody can know whether the second statement is true or false. Consequently, the answer would be marked 1 as shown below:

1 2 3 4 5

76. (A) A person who does not look you in the eye is likely to be dishonest.
(B) An interview with an emotionally upset client should be postponed until another day. 76

77 (A) Public records and documents are an optional source for verification of eligibility of applicants for public assistance by the visitor.
(B) Information is not included in the index of a social service exchange as to the treatment given to a registered individual. 77

78. (A) All property of a recipient of old age assistance is considered part of the recipient's estate in the Probate Court.
(B) The fact that an aged person is a recipient of a pension from some firm [illegible] industry [illegible] for 78

TOPICAL FILING

The five classifications in a subject file are as follows:

1. Accounting (includes accounts)
2. Administration (includes personnel)
3. Maintenance (includes equipment and supplies)
4. Sales
5. Transportation

Below is a series of topical sentences, names of catalogs, and the like which you are to classify according to the above five divisions. If the statement would be most logically classified under accounting, make a heavy mark on the ANSWER SHEET in space number 1; if it refers to administration, make the mark in space number 2, and so on.

EXAMPLES	Make a mark in space number
A copy of the building code	3
Regulations governing the wrapping of packages	5
A copy of corporation laws	2
Instructions to salesmen	4
Profit and loss statement	1

121. There is no express office in Fraser. 121
122. I am interested in learning more about your product. 122
123. I fear the chief account clerk will have to be discharged. 123
124. A list of freight stations. 124
125. [illegible]

PUNCTUATION

The following selection contains errors in punctuation. Read each sentence through first to get its meaning. Then correct the errors by crossing out needless punctuation, changing incorrect punctuation, and supplying omitted punctuation. Consider each of the following punctuation marks as one error:

period	.	semicolon	quotation mark
comma	,	colon	parenthesis) or (
hyphen	-	apostrophe	

When you have made all necessary corrections in punctuation, count the number of errors which occur in each line and make a heavy mark in the appropriate space on the ANSWER SHEET, thus:

Space number 1 if there is 1 error
Space number 2 if there are 2 or more errors
Space number 3 if there are no errors

151. Copyright laws have been in effect in the U S for more than 151
152. one hundred years; the first statute being passed May 31, 1790. 152
153. An author or owner, of unpublished material has a common-law 153
154. [illegible]

fangled ideas as scoring machines. Also, the State Technical Advisory Service of the Social Security Board was sufficiently interested in the mechanization of the selection process to encourage experimentation in that direction. Thus it happened that the Employment Board of the Pennsylvania Department of Public Assistance, a merit system agency whose history dates back only to the end of 1937, decided to score its examinations by machine.[3] Since then, the number of agencies with scoring machine installations has been steadily increasing.

Scoring Civil Service Tests with the I.B.M. Machine: Procedures Ensuring Necessary Accuracy

Insofar as the early reluctance to adopt machine scoring was based on skepticism concerning its accuracy, it was on firm ground, for when a civil service agency puts a score on a test paper, that score must be accurate. There is probably nothing more likely to undermine the prestige and public acceptance, if not the very existence under law, of a civil service commission than the frequent, or even infrequent, discovery of errors in its work. To the public, the exact nature of the procedures used by a commission in scoring its papers are relatively unimportant so long as they are honest and produce *correct* results.

In the early days of scoring civil service tests by machine it was thought that the procedures which were satisfactory for scoring educational achievement and similar tests would be equally satisfactory for scoring civil service examinations, if certain additional precautions to spot errors were taken. Such procedures have, however, been abandoned in almost every instance in favor of a system designed specifically to meet civil service commission requirements of accuracy. The procedure now in use by the majority of agencies gives results that are probably as accurate as it is possible to obtain while human beings operate the machines and practical considerations render it absurd to recheck other than borderline scores beyond

[3] Actually, the scoring was performed *for* the Employment Board by the Educational Records Bureau of New York City.

the point of finding more than one or two errors among thousands of scores.

This scoring system is suitable when the number of items answered correctly constitutes the written raw score, and depends for its accuracy upon recognition of the fact, as stated in the I.B.M. manual, that "the only truly accurate method of scoring is the one which takes into account every mark on the answer sheet which is intended as an answer, making allowance for questions answered more than once, and eliminating from the final score all stray marks not intended as answers by the examinee" (11). The five steps to this procedure are as follows:

1. Scanning papers for omissions and for items to which more than one answer has been indicated, and for the purpose of segregating sheets so poorly marked that they must be scored manually. A check mark is placed beside each omitted item and the number of items omitted is indicated in the box provided in the margin of the answer sheet. A horizontal line is drawn through each item answered more than once and the number of "surplus" answers indicated in the margin. (All marks on the answer sheet with the exception of those made by the examinee should, of course, be recorded with colored pencil.)
2. Scoring for rights on the machine.
3. Scoring for wrongs on the machine.
4. Totaling rights, wrongs, and omits (compensating for items answered more than once) and checking the total to see whether every item has been accounted for. If the total checks at this point, no further operations are necessary except manually scoring every 25th or 50th paper to provide a spot check of accuracy. (The additional precaution may be taken of manually scoring the answer sheets of all examinees whose scores range from two points below to one point above the passing point.)
5. Adjusting papers on which the total does not check in Step 4. When this is necessary, the paper goes to an adjuster whose job it is to determine the reason for the discrepancy and to correct it. Answer sheets which require such adjustment are then checked by a second adjuster to ensure accuracy. Answer sheets rejected in Step 1 as unsuitable for machine scoring are scored manually and checked by these adjusters or by others especially designated to perform this operation.

By means of the commoning key now available to scoring machine users, it is possible to perform a very useful screening operation in connection with Step 1, described above. This key may be inserted in the scoring rack between the sensing

and resistance units and the machine then adjusted so that the meter will indicate, for any given side of an answer sheet, the number of items attempted. For papers for which the machine adjusted in this way indicates "no omits," the task of scanning may be reduced to looking for items with multiple markings and papers which it is desirable to score manually.

A further refinement to this system may be wired into the machine so that the meter, instead of indicating the number of items attempted, will read the number omitted. This is accomplished by adjusting the circuit so that sufficient current will flow through the meter initially to indicate the total number of items in the test. Then, when an answer sheet is placed in the machine, the number of items attempted will automatically be subtracted from the initial reading, causing the meter to indicate the number of omissions.

Before leaving the subject of scoring, a few words of caution may be in order. The machine process in its present state of perfection is a big improvement over most other scoring techniques at present available for use by civil service jurisdictions. It has not, however, reached the state of refinement where it can be taken completely for granted that nothing will go wrong after the machine has been set up. Since every once in a while something does go wrong, it is necessary, to avoid later grief, not only to set up checks and controls but to require strict adherence to them on the part of all staff members charged with any scoring responsibility.

Other Machine Methods of Scoring

The I.B.M. scoring machine seems to offer, for ordinary civil service use, what appears to be the best all-around solution to many of the most annoying scoring problems confronting the medium or large size civil service agency. In addition, as will be noted later, this particular machine may be adapted for use in connection with several other examining tasks, including certain research projects that every civil service agency has in mind carrying through just as soon as the staff and the time are available.

There are, however, at least two additional scoring approaches classifiable under the head of "mechanical" that have been used to some extent by agencies conducting large numbers of examinations.

The first of these makes only partial use of a mechanical device—in this case a multilith, mimeograph or printing press—and is more accurately a technique for facilitating manual scoring than a procedure for scoring by machine. Separate answer sheets are used on which the examinee indicates his answers to multiple-choice or true-false items by checking appropriately numbered spaces. A multilith or mimeograph stencil is then prepared so that, when the answer sheets are run through the duplicating machine, a line connecting the correct answers will be printed over the response positions. When this has been done, it is possible for a scoring clerk to determine the number of correct answers simply by counting the number of responses marked in positions coinciding with the overprinted line.

This combination machine-manual method is considerably more rapid and accurate than manual scoring accomplished by placing a key alongside or over the answer sheet. By altering the procedure slightly it may also be used to advantage with completion-type questions. It requires, however, a skilled duplicating machine operator and the use of a duplicating machine capable of very accurate registration and extremely little spoilage. The multilith satisfies these requirements particularly well, and printing may, of course, be employed. On the other hand, the mimeograph appears to be less satisfactory.

A procedure, described by Iffert, Bloom, and Beum (6), for scoring multiple-choice tests by means of tabulating machines has apparently also been used with some success, although its chief value would appear to be in connection with the conduct of examining programs that are quite intimately tied in with research projects. When this particular method is employed, the examinee's answers are usually placed directly in the test booklet from which they are later punched into Hollerith cards and scored by successive runs through a sorter.

Once these cards are punched they are also available for research purposes and it is comparatively easy to conduct item analyses and compute correlations with them.

Scoring Graphic Rating Scales by Machine

Graphic rating scales may be, and have been, scored by the I.B.M. scoring machine. "By using the aggregate weighting unit of the machine it is possible to obtain the aggregate weighted average of as many as 30 variables, each varying in size from 1 to 100 and (in groups of three) weighted from 0 to 20" (10).

In utilizing this feature of the machine the rating scale is usually designed so that it is necessary to draw a horizontal line for each characteristic rated. The length of each such line determines the score for that particular characteristic, and the weighted total for all characteristics is indicated on the meter in the same fashion as any other score. The horizontal lines should, of course, be drawn with a special pencil, and may be made by the rater himself or be drawn in later by a clerk.[4] When the latter plan is used, the rater checks (with a colored pencil) the point on each line that represents the rating he wishes to assign and the clerk simply draws a line from the origin (left) to each check mark. Graphic scales of this type may be used in connection with oral interviews, service ratings, or performance test ratings.

Scoring Training and Experience by Machine

Many civil service commissions include a quantitative rating of training and experience in the test battery for a majority of the classes of positions for which they conduct examinations. It now appears quite likely that a considerable portion of the computational work connected with the use of the type of training and experience rating scale (14) employed, in one

[4] In several informal studies conducted by an agency which formerly used this type of rating sheet in large quantities it was found that where raters made check marks only, it was apparently faster to score scales of this kind manually (by having a clerk place a stencil over the rating sheet and add the numbers on a comptometer) than to go to the trouble of drawing a line for each characteristic before running the sheets through the scoring machine.

form or another, by numerous agencies throughout the country may be performed by machine.

This possibility was brought closer to realization recently with the development[5] of a tentative form of machine-scorable training and experience rating scale which, while it still remains to be tried in an actual test situation, looks very much as though it will not only work but possibly be at least a partial or preliminary answer to the mechanization of this phase of the selection process.

This adaptation of the machine makes use of the principle of the previously mentioned commoning key. In marking the scale the rater simply blackens each space that corresponds to a type of training or experience possessed by the examinee whose application is under consideration. As an example of the possibilities of this approach, provision has been made in the initially developed form of the scale for recognition of up to 15 years—in six-month steps—of each of four levels of related experience, three levels of related under-graduate and graduate study, and the possession of academic degrees.

Scoring Service Ratings by Machine

Present service rating instruments take various forms, many of which may be scored by machine. Before deciding to adopt such a procedure, however, the numerous factors involved should be given careful consideration, and the decision based upon the extent to which machine scoring will contribute to the economy, speed, and all-around efficiency with which this particular phase of the program may be administered.

Service rating scales of the graphic type may be scored by setting them up to utilize the aggregate weighting feature of the machine. In addition various adaptations of the graphic approach may be employed which, depending upon the particular situation at hand, may be scored by using either the ordinary answer key form alone or in conjunction with the more recently available commoning key, with the latter set-up offering considerably the greater possibilities. Service rating

[5] This machine-scorable scale was the outgrowth of a discussion participated in by E. C. Schroedel, G. C. Sloughter, J. H. Pockrass, and S. W. Koran.

forms of the check-list variety may also rather easily be adapted to scoring by machine.

Item Analysis With the Scoring Machine

A recently developed attachment to the scoring machine is the graphic item counter, which is available as optional equipment. This device consists of a plugboard having a plugging position for each of the 750 response positions on the standard answer sheet and for each of 90 counters. By means of plugwires, any response position may be connected to any counter. When the appropriate response positions and counters have been wired together, the plugboard is inserted into the machine in the position normally occupied by the scoring rack.

Using this attachment it is possible to secure, in a single run through the machine, a graphic count of the marks placed in up to 90 response positions on 100 answer sheets. If more than 100 sheets are involved in the study, a separate graphic count must be made for each group of 100 sheets. If more than 90 response positions are to be analyzed in a given test, the plugboard must be rewired and the sheets run through again for the additional responses. Thus, if in a given analysis, it is desired to determine the number of individuals who correctly answered each of the 150 five-choice items on a single side of an answer sheet and the population of the study is 175, it is necessary to run the 175 sheets through the machine twice, making separate graphic counts of the first 100 and the last 75 on each of the two runs. Items 1 to 90, inclusive, may be analyzed on the first run, and the remaining 60 items (91 to 150, inclusive) on the second run.

If the item analysis is of the variety that requires information concerning the examinees' selection of each of the five possible responses to the 150 items, the sheets will have to be run through the machine nine times to obtain this information for each of the 750 possible responses. Whether the items will be analyzed to the extent of determining the number of examinees selecting each possible response or be confined to determining the number of examinees selecting the correct answer

will, of course, depend upon the use or uses for which the data are intended. The speed of operation of the machine equipped with the item analysis unit has been reported as ranging from 400 to 500 papers per hour when 90 responses are analyzed on each paper. This is considerably faster than any clerk can perform the job manually.

Computing Reliability Coefficients, Standard Error of Measurement and Intercorrelations with the Scoring Machine

During the past few years several techniques have been developed for using the test scoring machine to facilitate the computation of such useful measures as intercorrelations, reliability coefficients, and the standard error of measurement. While it is beyond the scope of this presentation to go into the derivation of the formulas that have been developed or to describe the procedures at any length, mention will be made of a few of the more important of these applications in order to illustrate the variety of the scoring machine's uses in connection with examination research.

One kind of research, that of investigation into the validity of individual items by means of the technique of item analysis, has already been mentioned. Hoyt (5) recently described a method of computing test reliability which makes further use of some of the data obtained when such an item analysis is performed. The procedure which Hoyt suggests was developed as a practical and simplified application of Richardson and Kuder's (9, 15) "method of rational equivalence," which produces a coefficient of reliability that in certain respects appears to be superior to that obtained by using the split-half correlation method with the Spearman-Brown formula. No data beyond those obtained when a test is scored and an item-analysis performed are required for substitution into the following formula (5):

$$r_{tt} = \frac{n}{n - 1} \cdot \frac{kSs + Si - T(T + k)}{kSs - T^2}$$

in which r_{tt} is the reliability of the test, n is the number of items in the test, k is the number of subjects taking the test,

T is the sum of the scores obtained by all the subjects, *Ss* is the sum of the squares of the scores obtained by the subjects, and *Si* is the sum of the squares of the number of correct responses to each item.

Although the Kuder-Richardson technique has numerous advantages which make its wider adoption quite likely, many investigators in the field of civil service examinations may have to continue to obtain most of their reliability coefficients by means of the split-half method used in conjunction with the Spearman-Brown formula. This, at any rate, will probably continue to be the situation unless item analysis data are available for substitution into a formula such as the one above or the simpler formula presented by Kuder and Richardson[6] is not appropriate in the specific situation.

When it is known at the outset of scoring a given test that a split-half reliability coefficient will be required, it is possible to prepare the scoring matrices so that separate scores for the odd-numbered and for the even-numbered items will be obtained when the answer sheets are run through the machine for the first time. This may be accomplished by keying the odd-numbered items in the usual fashion (that is, as rights), and the even-numbered items as wrongs (that is, preparing the scoring matrix so that even-numbered items answered correctly will be indicated when the selector switch is in the wrongs position). To secure the total rights score when the machine is set up in this way, the selector switch need only be moved to the R + W position. When the selector switch is moved to the R position the meter will indicate the number of odd-numbered items answered correctly, and when it is in the W position, the number of even-numbered items answered correctly.

Far from being extra work, this procedure possesses the

[6] It should be noted that Kuder and Richardson (9) have derived a simpler formula (No. 21 in the article referred to) which can be computed in two or three minutes, given the number of items in the test, the average score, and the standard deviation of the scores. It gives a slight underestimate of the true reliability of a test. For the more reliable tests the estimates obtained by this formula are usually from .01 to .03 less than those obtained by use of the more rigorous formulas presented.

advantage of providing a useful check on the total score. As only half of the sets of contacts are connected to the meter when the switch is in either the R position or the W position, the effect of stray marks on the reading is frequently minimized to the point where, on some papers, the total of the separate R and W readings (added without the use of the scoring machine) may provide a more accurate score than the R + W (total rights) reading taken alone. The reason for this is, of course, that while the effect of stray marks and poor erasures may be sufficient to influence the score when 150 sets of contacts are in the circuit, their effect may not be noticeable when only 75 sets of contacts are involved at a time.[7]

It might be well, at this point, to call attention to the fact that when the "rights plus wrongs plus omits" scoring procedure described earlier in this paper is employed with a single machine set up to read rights and wrongs on a single insertion, it is not possible to secure the split-half scores at the same time in the fashion just suggested. This problem may, however, be solved by reading the rights on one run through the machine and the wrongs on a subsequent run.

Several formulas are available for determining split-half reliability. In place of the orthodox product-moment formula, some investigators prefer, under certain circumstances, a formula which requires the substitution of values obtained from the odd scores and total right scores only. As described by Mosier (13) this formula takes the form:

$$r_{oe} = \frac{r_{ot}\sigma_t - \sigma_o}{\sqrt{\sigma_t^2 + \sigma_o^2 - 2r_{ot}\sigma_o\sigma_t}}.$$

[7] The problem presented by the necessity for manually scoring considerable numbers of answer sheets because the effect of stray marks causes the total of rights, wrongs, and omits to differ from the total number of items is worthy of attention. The kind of solution suggested above in connection with the split-half scoring of rights may, if necessary, be adopted as a regular practice in scoring wrongs as well as rights. Because of the considerably larger answer sheet area exposed to "live" contacts when the wrongs score is being determined, such readings are usually affected by stray marks and poor erasures to a greater degree that rights scores. When split-half scores are not required, the answer sheet area may, of course, be divided into two or three sections by punching appropriate field selection holes so that a separate reading may be obtained for each area.

Still another approach—and one which has been gaining considerable favor recently with scoring machine users—makes use of the fact that once the machine has been set up to indicate odds and evens, the difference between these values may be obtained by simply turning the selector switch to the R-W position. The standard deviation of the difference scores thus obtained for a given test will, as has been pointed out by Rulon (16), equal the standard error of measurement of the scores in that group. A split-half reliability coefficient may then be obtained by substituting in the following simple formula (2):

$$r_{11} = 1 - \frac{\sigma_o^2}{\sigma_w^2}$$

in which r_{11} is the reliability of the test, σ_o is the standard deviation of the difference between the odd and even scores, and σ_w is the standard deviation of the ordinary test scores including both odds and evens.

A method making it possible to use the scoring machine for calculating tables of intercorrelations has been developed by Kuder (8) but has apparently not been very widely used, perhaps because of the laborious clerical work involved in preparing the coded answer sheets required for the computations. The work thus involved has lately been reduced, however, in a revised and simplified procedure suitable for studies involving up to 150 cases. The Kuder technique is an adaptation of the Royer-Toops method of obtaining correlations from Hollerith cards on which geometric codes of scores for each variable have been punched. Kuder's approach has been to use an answer sheet and stencil for each code, but he does not recommend substituting the scoring machine for Hollerith equipment when the latter is easily available (8).

Kuder has also pointed out that the scoring machine is excellently suited for obtaining tetrachoric coefficients of correlation and that the procedure for doing so is relatively simple. Since in tetrachoric correlation each variable is divided into a dichotomy instead of into class intervals, the amount of "coding" and clerical work involved is reduced to a minimum.

Tabulating Equipment

It has been possible to note, during the past few years, a considerable increase in the number of civil service agencies that have adopted mechanical procedures for handling, in addition to scoring, such other operations as assigning candidates to the written, oral, and performance tests; computing grades; notifying candidates of their eligibility and ineligibility; establishing registers of eligibles and lists of candidates who did not qualify; certifying names from registers; maintaining miscellaneous personnel and payroll records; and aiding in the conduct of research (3, 4). Let us examine briefly some of the ways in which Hollerith equipment may be used to facilitate the conduct of some of the operations connected with the processing of examinations, keeping in mind, however, the unlikelihood that a machine installation would prove particularly economical for conducting these specific operations to the exclusion of the numerous related tasks just enumerated.[9]

One of the more important jobs which it is possible to perform successfully with Hollerith cards is that of converting, weighting, and combining examiners' scores on the various components of the examination, transmuting the results into final grades, adding veterans' credit and other bonuses, and listing, in order of final grade, the names of those who have qualified.

The tabulating card used for this purpose includes fields which provide columns into which may be punched such data as the following, depending upon the needs of the given situation: identification and file numbers; class of position; written, training-experience, oral, performance, and service rating raw and converted (or weighted) scores; total converted score, final grade, veteran's credit, rank, *et cetera*. Written raw scores and identifying data are punched into what are called "detail" cards with the electric key punch and are verified by means of the mechanical verifier. These cards are then ar-

[9] An exception to this might be the situation in which certain equipment of another agency is available for part-time use so that the only additional machines required are a key punch and verifier, and possibly a sorter.

ranged in order of identification number by means of the horizontal sorter. Raw scores on each subsequent part of the examination battery are usually first punched into "scratch" cards which, after being verified, are sorted according to identification number in the same fashion as the detail cards. When this has been done, the data on the scratch cards are transferred to appropriate columns of the detail cards through use of the automatic reproducing punch (3).

After the raw scores have been punched into the detail cards, it is usually necessary to convert them into whatever variety of transmuted score the agency uses for the purpose of assigning the announced weight to each component of the examination. The raw score data which ordinarily serve as the basis for computing the conversion tables are easily secured from the cards by sorting them by raw score and running them through a numeric tabulator.

Several methods of transferring the transmuted scores to the detail cards may be used: Conversion tables may be prepared for use by key punch operators who determine the converted score corresponding to each raw score and punch that figure into the detail card (4). A second method that may be employed calls for using the automatic multiplying punch for the purpose of multiplying the raw score by some constant (for example, the reciprocal of the number of questions in the written test, if a percentage is desired). A third method makes use of prepunched master cards each of which contains a possible raw score and its corresponding conversion. When using this arrangement, both the master cards and the detail cards are sorted by raw score and run through an automatic reproducing punch which transfers the converted scores from the master cards to the detail cards at a high rate of speed (3). When the transmuted scores for all components of the examination have been entered, they may be totalled and the sum punched into the appropriate columns of the detail card. If this total requires further conversion, the process involved is identical to that of transmuting individual raw scores and may be carried out in any one of the three ways mentioned.

In addition to serving the purposes for which they were designed, the punched cards used in arriving at the examinee's final grade and register position are also available for numerous research uses. Agencies to which the use of a Hollerith installation is available are in the fortunate position of being able to perform many of the research jobs discussed as possible with the scoring machine and, in addition, to make use of the amazing flexibility of the punched card method to conduct types of research which, because of the amount of clerical and statistical work sometimes involved, are all but impracticable when attempted without such aid.

REFERENCES

1. Dunlap, Jack W. "Problems Arising from the Use of a Separate Answer Sheet," *Journal of Psychology,* X (1940), 3-48.
2. Flanagan, John C. "Note on Calculating the Standard Error of Measurement and Reliability Coefficients with the Test Scoring Machine," *Journal of Applied Psychology,* XXIII (1939), 529.
3. Hawthorne, Joseph W. and Morse, Muriel. *Business Machines in Public Personnel Administration,* Los Angeles City Civil Service Commission, 1940, 43 pp.
4. Horchow, Reuben. *Machines in Civil Service Recruitment.* Chicago: Civil Service Assembly of the U. S. and Canada, Pamphlet No. 14, 1939, 43 pp.
5. Hoyt, C. J. "Note on a Simplified Method of Computing Test Reliability," *Educational and Psychological Measurement,* I (1941), 93-95.
6. Iffert, R. E., Bloom, B. S., and Beum, C. O. *Another Test-Scoring Procedure: A Method of Scoring Short Tests on the Hollerith Sorter.* Columbus: Ohio College Association Bulletin, No. 118, Mimeographed, February 1940, 7 pp.
7. Koran, Sidney W. "Adapting Tests to Machine Scoring," *Journal of Applied Psychology,* XXIII (1939), 709-719.
8. Kuder, G. Frederic. "Use of the International Scoring Machine for the Rapid Calculation of Tables of Intercorrelations," *Journal of Applied Psychology,* XXII (1938), 587-596.
9. Kuder, G. F., and Richardson, M. W. "The Theory of the Estimation of Test Reliability," *Psychometrika,* II (1937), 151-160.
10. *Machine Method of Scoring and Analyzing Examinations.* New York: International Business Machines Corporation, undated, 14 pp.
11. *Machine Methods of Test Scoring: Manual of Procedures.* New York: International Business Machines Corporation, 1940, 7 pp.
12. *Manual of Instruction for the International Test Scoring Machine.* New York: International Business Machines Corporation, 1939, 20 pp.
13. Mosier, Charles I. "A Short Cut in the Estimation of Split-Half Coefficients," *Educational and Psychological Measurement,* I (1941), 407-408.
14. Pockrass, Jack H. "Rating Training and Experience in Merit System Selection," *Public Personnel Review,* II (1941), 211-222.
15. Richardson, M. W., and Kuder, G. F. "The Calculation of Test Reliability Coefficients Based on the Method of Rational Equivalence," *Journal of Educational Psychology,* XL (1939), 681-687.
16. Rulon, Phillip J. "A Simplified Procedure for Determining the Reliability of a Test by Split Halves," *Harvard Educational Review,* IX (1939), 99-103.

PREDICTIVE VALUE OF CERTAIN "LAW APTITUDE" TESTS[1]

E. L. WELKER and T. W. HARRELL
University of Illinois

THIS PAPER REPORTS the second of a series of studies analyzing the abilities necessary for success in law school. An earlier study[2] showed that pre-law grades from one school, the University of Illinois, correlated higher with law grades than did the Ferson-Stoddard *Law Aptitude Examination.* Combining the test and pre-law grades did not significantly improve the prediction. The homogeneous parts of the law aptitude test were correlated separately with law grades and showed that the memory case questions gave an unquestionably insignificant correlation. This result is interesting since the memory material seems to represent a popular stereotype of what a law student has to do.

Several other investigators have reported comparisons between test scores and law-school grades, but apparently no one has previously reported a detailed attempt to analyze the relation between separate law course grades and part scores of "law aptitude tests." The ultimate aim of these studies is of course to discover tests that will lead to the more valid prediction of law school success.

The variables included in this study are listed in Table 1. It will be noted that the tests used were the homogeneous parts of the Ferson-Stoddard *Law Aptitude Examination,* in addition to the homogenous parts of other selected tests—the Yale *Legal Aptitude Test,* the American Council on Educa-

[1]This study was made possible through the generous cooperation of Dean Albert J. Harno, University of Illinois College of Law.

[2]T. W. Harrell, "Predicting Success of Law School Students." *American Law School Review.* IX (1939), 290-202.

tion *Psychological Examination,* and the comprehension and speed tests of the *Minnesota Reading Examination.* Some of the pre-law grades (variable 26) were approximated where students attended a school other than Illinois. Average law grades for the first semester as well as course grades for the five first-semester courses were used as criteria.

TABLE 1. NAMES AND DESCRIPTIONS OF VARIABLES

Variable Number	Description	Name
1	Interpretive case	Ferson-Stoddard Law Aptitude Exam. Part 2-A
2	Completion	Ferson-Stoddard Law Aptitude Exam. Part 2-B
3	Relevant facts	Ferson-Stoddard Law Aptitude Exam. Part 2-C
4	Logical inferences	Ferson-Stoddard Law Aptitude Exam. Part 3
5	Matching	Ferson-Stoddard Law Aptitude Exam. Part 4
6	Memory case	Ferson-Stoddard Law Aptitude Exam. Part 1-B
7	Arithmetic case	ACE Psychological Examination (1938)
8	Pattern analogies	ACE Psychological Examination (1938)
10	Completion	ACE Psychological Examination (1938)
11	Artificial language	ACE Psychological Examination (1938)
11	Artificial language	ACE Psychological Examination (1938)
12	Same-opposite	ACE Psychological Examination (1938)
13	Reading speed	Minnesota Reading Examinations
14	Reading comprehension	Minnesota Reading Examinations
19	Word relations	Yale Legal Aptitude Test Group I
20	Opposites	Yale Legal Aptitude Test Group II
21	Word analogies	Yale Legal Aptitude Test Group III
22	Logical inferences	Yale Legal Aptitude Test Group IV
23	Memory case	Yale Legal Aptitude Test Group V
24	Interpretive case	Yale Legal Aptitude Test Group VI
25	Definitions	Yale Legal Aptitude Test Group VII
26	Pre-Law Grades, including those approximated from other schools	
27	Average First-Semester Law Grades, Univ. of Illinois College of Law	
28	Course Grades in Contracts	First Semester, Univ. of Illinois Col. of Law
29	Course Grades in Torts	First Semester, Univ. of Illinois Col. of Law
30	Course Grades in Remedies	First Semester, Univ. of Illinois Col. of Law
31	Course Grades in Criminal Law	First Semester, Univ. of Illinois Col. of Law
32	Course Grades in Possessory Estates	First Semester, Univ. of Illinois Col. of Law

The subjects were 133 male Law College freshmen at the University of Illinois. Seventy-eight of these entered in the fall of 1938 and 55 in the fall of 1939. The means of the two groups on both test scores and grades appeared similar enough to justify combining the data for the two years into one study.

The product moment coefficients of correlation between each of 21 test scores and average first-semester law grades are shown in Table 2. Insignificant correlations, i.e., those

less than .17, for which the chances that such a coefficient of correlation will occur in an uncorrelated population are more than 5 in 100, are omitted. Barely significant coefficients, i.e., those between .17 and .22, where the chances are more than 1 in 100 that such a coefficient will occur in an uncorrelated population, are in parentheses. No correction has been made for attenuation or the unreliability of the variables.

TABLE 2

PRODUCT MOMENT COEFFICIENTS OF CORRELATION WITH FIRST-SEMESTER LAW GRADES N — 133

Variable Number	Correlation with First Semester Law Grades
1	—
2	—
3	(.17)
4	.28
5	.31
6	—
7	.23
8	(.17)
9	.24
10	.28
11	.31
12	—
13	—
14	.25
19	.25
20	.39
21	.33
22	.30
23	(.19)
24	—
25	—
26	.49

Note: Insignificant coefficients are omitted and barely significant ones are parenthesized.

In evaluating some of the parts of the Yale *Legal Aptitude Test* it should be noted that the scores reported do not represent separate sections with individual time limits. The test is made up of three parts which are separately timed. The parts are not homogenous as to the type of item used. The first and third parts are composed of four item types: Word Relation, Word Opposites, Word Analogies, and Logical Inferences. These are arranged in cycle-omnibus form with 10 items of the same type together. The total number

of items of each type is 40. The second part is made up of three additional kinds of questions. First are 20 memory items dealing with a case that was presented at the beginning of the test—before Part 1. Next are 40 items of the Interpretive Case variety. Finally there are 20 Definition questions.

Mr. W. E. Kline of the Yale Personnel Bureau writes that the Interpretive Case and Definition items have consistently yielded only low correlations with grades. This result is explained by the fact that these types of questions do not appear early enough in a timed section for reliable scores to result. Consequently a new form of the Yale test is being put together. "It contains seven sub-tests, each of which is homogeneous and individually timed."

Tests correlating clearly significantly with law grades, as shown in Table 2, are, in order of their coefficients from high to low: Yale Opposites, Yale Word Analogies, Ferson-Stoddard Matching, ACE Artificial Language, Yale Logical Inferences, Ferson-Stoddard Logical Inferences, ACE Completion, Minnesota Paragraph Reading Comprehension, Yale Word Relations, ACE Number Series, ACE Arithmetic. Tests adjacent in order seldom if ever have coefficients that are significantly different.

It is recognized that a completely thorough understanding of the interrelations of the variables calls for a factor analysis. Such a study is planned. All intercorrelations have been computed.

Tests which correlated barely significantly with law grades, as shown in Table 2, are: Yale Memory Case, Ferson-Stoddard Relevant Facts, and ACE Pattern Analogies.

The following tests did not correlate significantly with the first-semester mean: Ferson-Stoddard Interpretive Case, Ferson-Stoddard Analogous Case, Ferson-Stoddard Memory Case, ACE Same-Opposite, Minnesota Reading Speed, Yale Interpretive Case, and Yale Definitions.

None of the correlations is as high as .40. The Yale test correlates slightly higher than any other test total. Some of

the American Council sub-tests and the Minnesota Reading Comprehension correlate significantly, while some of the so-called law aptitude sub-tests do not.

It was mentioned above that the previous study showed that the memory case questions in the Ferson-Stoddard test correlated insignificantly with law grades. This result is confirmed here, but the Memory Case in the Yale examination does give a barely significant correlation. Mr. Kline writes that the memory questions correlated .33 with first-year grades of the Yale Law freshmen of 1940.

Pre-law grades correlated .49 with first-semester law grades. This coefficient is higher than any with test scores, but considerably lower than that reported in the previous paper. One explanation for the lower coefficient is the lessened accuracy of the present pre-law grades. These include grades at schools other than Illinois plus those at Illinois. Previously only Illinois pre-law grades were included. Where grades from different schools are combined, it seems unlikely that the result will be as reliable a test of values as those from one school, due to differences in grading systems. Another reason for the decreased correlation between law grades and pre-law grades is that the present group is more homogeneous for pre-law grades. This situation was occasioned by raising the requirement for entrance for Illinois students having only 3 years' credits from a grade-point average of 3.0 to 3.25.

The product moment coefficients of correlation between each of five law grades and the 21 test variables are shown in Table 3. Again, insignificant coefficients have been omitted, and barely significant ones parenthesized as in Table 2. Variable 30, Remedies, correlated significantly with 12 test scores; variable 28, Contracts, with nine; variable 29, Torts, with nine; variable 32, Possessory Estates, with four; and variable 31, Criminal Law, with only two. Legal aptitude tests measure more nearly what is required to master Remedies, Torts, and Contracts, than they measure what is required to understand Criminal Law and Possessory Estates.

The significance of these differences has not been tested. Part of the differences could be otherwise explained if the reliability of the grades varied markedly from one course to another, but their reliabilities are unknown. Some thought has been given to estimating the reliabilities using the Kuder-Richardson method, since it does not demand split-half scores. This has not been done because the number of items represented by the law grades is nonexistent. Some conversation with lawyers suggests that Remedies does demand more reasoning than does Criminal Law, which requires greater memorization.

TABLE 3.

PRODUCT MOMENT COEFFICIENTS OF CORRELATION BETWEEN EACH OF 5 LAW GRADES AND 21 TEST SCORES.

Variable Number	28	29	30	31	32
1	—	—	—	—	—
2	—	—	—	—	—
3	—	(.17)	(.20)	—	(.19)
4	.32	.24	.31	—	(.22)
5	.28	.31	.34	(.19)	.25
6	—	—	—	—	—
7	.23	(.19)	.25	—	.24
8	—	(.22)	.23	—	—
9	(.22)	(.19)	.30	—	(.18)
10	.26	.29	.30	.26	(.17)
11	.37	.26	.41	(.20)	(.22)
12	—	—	(.18)	—	—
13	—	—	—	—	—
14	(.21)	.24	.26	—	—
19	(.21)	.24	.35	—	—
20	.39	.35	.47	.26	.30
21	.29	.32	.44	(.22)	.24
22	.25	.27	.39	(.22)	—
23	.25	(.22)	—	—	(.19)
24	—	—	(.21)	—	—
25	—	—	(.22)	—	—

Note: Insignificant coefficients are omitted and barely significant ones are parenthesized.

It will be noted that three of the correlations with Remedies are higher than any of those with the semester means. The differences are scarcely reliable.

It can be tentatively concluded that while no legal aptitude test correlated as high with law grades as do pre-law grades, the most predictive tests are those that call for reasoning rather than memory. The reasoning tests may use

words or numbers for symbols, but there seems to be an advantage for the former, as might be expected.

Each of the two legal aptitude tests correlates higher with pre-law grades than with law grades. This difference might be explained if the pre-law grades are more reliable than the law grades. The authors have not been able to determine the reliability of either. The law grades might be expected to be more reliable from the fact that the law course is more homogeneous than is the varied pre-legal curriculum. On the other hand, grades based on 6 to 8 semesters of pre-law work, because of the increased reliability with additional length, would be expected to be more reliable than grades from a single semester of law.

Since the two so-called "legal aptitude" tests correlate lower with law grades than with other college grades and since several tests that are not called "legal aptitude" correlate higher with law grades than several that are putative measures of law-school success, the question is raised as to the possible existence of a factor or factors of legal aptitude. The factor analysis of these data may contribute a clearer answer to the question.

AN EXPLORATORY STUDY OF SOCIAL GUIDANCE AT THE COLLEGE LEVEL[1]

MARGARET GLOCKLER ALDRICH
University of Minnesota

WITHIN THE LAST 10 years there has been an increasing emphasis on guidance at the college level. This movement is important, but it is significant that the personnel workers in institutions of higher education have been concerned almost exclusively with educational and vocational problems. In some cases, however, college authorities have come to realize that there are certain social problems and adjustments which should be considered. Many college programs fail to provide social stimulation and opportunity for participation. Extra-curricular activities have developed on college campuses to fill this need. Most of these activities have been developed on the basis of student initiative in spite of, rather than because of, faculty approval.

An interest in the development of social adjustment in colleges led to the following experimental evaluation of social guidance at the college level.

Naturally, the valuation of guidance has lagged far behind the development of guidance techniques. There have been several attempts to determine the effects of diagnosis and treatment of educational and vocational problems [Beaumont

[1]This study was undertaken at the suggestion of Professor D. G. Paterson of the University of Minnesota. His advice and interest made its completion possible. Dean E. G. Williamson, Dr. John Darley, and the counselors of the University of Minnesota Testing Bureau, as well as various extra-curricular organizations on that campus, made possible the execution of the problem.

(1), Wrenn (10), and Williamson and Bordin (8)]. These studies indicate the possibilities in this field. However, there has been no such study of social guidance, even though several workers have recognized the need for evaluating this type of guidance [Tuttle (7), Livingood (5)]. Others [Burke (3), Mallay (6), and Williamson and Darley (9)] have attempted to distinguish the socially "well" adjusted from the socially "poorly" adjusted. But in the psychological literature of the last five years there is no report of a study of the effect of any particular controlled factor on social adjustment.

Several investigations at the University of Minnesota have demonstrated the need for further concern with the social and extra-curricular program [Chapin (4), Brown (2)]. The more extensive of these was that by Brown in 1934. She found that one third of the students spend no time or money in activity participation and concluded that "students most needing social contacts were those who profited least from the opportunities offered." (2:263)

These conclusions might be interpreted to mean that there is little hope of better adjusting asocial students. Another possibility is that if these students who did not participate were given an intensive well-directed program of participation, they might become better adjusted socially. Is it possible to lead the asocial to activities and find any changes in their social interests and attitudes?

The essential plan of this research was to expose students to certain social influences and measure any changes resulting from the contacts formed. To be of value, it was necessary that these influences be normal extra-curricular and counseling activities available to all college students. It was also necessary that the control group technique be used to determine what would occur without these special influences. This need led first to a consideration of possible methods for measuring changes that occur in the social adjustment of college students. To make the group as homogenous as possible, it seemed advisable to limit the study to freshman girls. Since all the

girls were to be treated as a part of a normal counseling program, it was further necessary to study only girls who had gone through the University Testing Bureau. This Bureau is a counseling agency set up by the University as a personnel service open to all students. Testing Bureau cases are given a rather extensive testing program including a series of personality scales. During the summer of 1939, 198 freshman girls came to the Bureau for guidance prior to registration in the University. From this group the experimental group was further selected by the requirement that the research be done on asocial girls as indicated by personality test scores and activity records.

These conditions help to explain why the general problem of the effect of social guidance becomes quite specific, i.e., what is the extent of change, if any, in the measured social adjustments and activity records of "under-socialized" University Testing Bureau freshman girls following counseling on social problems and directed participation in extra-curricular and social activities?

The first step in the attack on this problem was the selection of the sample group. The case records of the 198 Testing Bureau cases were read and a record kept of high school scholarship percentile rank, raw score and percentile rank on the American Council *Psychological Examination,* the *Cooperative English Test,* the *Minnesota Inventory of Social Attitudes*—Forms P and B, the Bell *Adjustment Inventory*—Social, and the *Rundquist-Sletto Inferiority Scale.* In addition, each girl's group and individual activities listed on the Individual Record Form of the Testing Bureau were recorded. These two sections are given in the form of a check list on which the subject is asked to indicate those activities "in which you engage frequently." The group activities include team sports, clubs, church organizations, and group parties, while the individual activities are things done alone or with a single other individual, such as sewing, reading, and tennis.

Those girls who had scores in the lower one half of two of the three distributions of Social Preferences, Social Be-

havior (*Minnesota Inventory of Social Attitudes*—Forms P and B), and group activities were selected as the sample group. These three measures were assumed to give an indication of the preferences, behavior, and previous interest in social activities. It should be noted that in order to get a sample of any size the levels had to be fairly high, running up to the median or even higher. This selection from the 198 Testing Bureau cases yielded a sample of 79 freshman girls. This group was then divided into two random samples of 40 and 39, which became the experimental and control groups. The remaining 119 Testing Bureau cases were used for purposes of comparison.

The treatment of these two groups must be emphasized here since this is the crux of the method. The control group of 39 cases was in no way influenced by this study. These girls were handled in the customary manner by the Testing Bureau. Following a preliminary interview and testing, each girl was assigned to one of the five counselors in the Bureau. The usual counseling interview is chiefly concerned with educational and vocational problems. If the social aspects are of importance, however, they may also be considered. No generalizations can be made concerning the social counseling of the control group except that they were exposed to the "normal" counseling program which might include some social guidance.

The treatment of the experimental group, however, went further in giving all of the members of this group an opportunity to participate in social activities. Nine girls in this group as well as 11 in the control group failed to complete the counseling and retesting program. Seven girls in this group were already participating, and the counselors merely discussed their social interests with them. Four were untreated because the counselors felt that academic activities should take all of their time if they were to continue in school. The remaining 20 were interviewed by the counselors with a special emphasis on social adjustment. Each interview took place at the end of the first quarter of the school year. The investigator consulted with the counselors concerning the activities

which might appeal to each girl, but the interview was very much an individual affair. Following this contact, the investigator attempted to carry out the suggestions of the counselors by personally introducing the girl to those activities in which she expressed an interest. Active participation was facilitated in every way possible. At the same time every precaution was taken to make the social program normal. Introductions were made to various campus organizations which had been informed that the Testing Bureau had appointed a special counselor to act as liaison officer between the Bureau and the organization. The extra-curricular organization heads did not know that this was in any way a research project, and there is every reason to believe that they gave these girls attention similar to that given any girls recommended by a campus agency. The experimental and control groups both had some counseling, but the experimental group had more than the control group.

At the end of the school year both the experimental and the control groups were retested. After three notices all but 20 girls responded; of the 79 originally selected for study, 31 experimental and 28 control subjects were retested. The tests which were given again were the Bell, the Rundquist-Sletto, the Social Preferences, and the Social Behavior. Also included was an Activity Record covering the freshman year. The tests were given under conditions identical with those of the original testing.

It is important that some consideration be given to the original nature of the control and experimental groups. Using the common *t* test for the significance of the difference between two means,[2] the sample group (the 59 cases who were retested) does not differ from the rest of the Testing Bureau cases (the Freshman women Bureau cases which were read but not selected for this study) in mean American Council on Education test score or the *Cooperative English Test* score. It does have a significantly higher mean score than frequently

[2] *t*=the difference in means divided by the standard error of that difference.

used norm groups. As would be expected, the experimental and control groups are significantly lower in mean score on the Social Behavior and Preference scales when compared with the rest of the Bureau group, but they do not differ in mean Social Behavior score from college Freshman women norm groups as determined from the norms given by the authors of the test. The difference between the mean Social Preference scores for the sample group and the same norm group is just significant (with the sample group lower). The experimental and control groups differ on the Bell and Rundquist-Sletto from the rest of the Bureau group but not from comparable norm groups. These facts lead to the conclusion that although the experimental and control groups are socially "poorly" adjusted when compared with the rest of the Testing Bureau group, they are not clearly different on personality measures from the comparable norm groups. Thus this study was not confined to a group of extreme deviates in personality scores.

Although the control group has a slightly higher median high school percentile rank than the experimental group, the two groups are remarkably similar in original testing on the six objective measures, and all indicated group and individual activities. It is, therefore, safe to assume that any significant differences on retesting may be attributed to differential treatment. It was also found that the 20 cases who did not appear for retesting did not differ significantly on original testing from the rest of the sample group.

The differences on retesting between the experimental and control groups provide the basis for an estimate of the success of social guidance. It would be desirable to obtain some estimate of the amount of guidance and relate this to the amount of participation, although this was not done here. A simple comparison was made of mean gains. These results and those from the Activity Record, using *t* tests where possible, indicate that:

1. There was a significant mean gain made by both the experimental and control groups on retesting on the Rundquist-Sletto Inferiority, the Social Preferences, and the Social Be-

havior scales. The control group did not gain on the average on retaking the Bell Social scale, while the experimental group did gain.

2. A comparison of the mean gains made on retesting after 9 to 11 months on these measures by the experimental and control groups shows that on all except the Social Preference scale the experimental group gained significantly more (see Table 1).

TABLE 1

MEAN GAINS MADE ON THE FOUR PERSONALITY MEASURES BY THE EXPERIMENTAL AND CONTROL GROUPS ON RETESTING

	Experimental			Control				
Measure	N	Mean Gain	S.D.	N	Mean Gain	S.D.	t	Pt
Social Beh.	30	4.40	10.12	28	2.21	12.83	3.65	<.01*
Social Pref.	30	6.20	13.00	28	6.18	15.47	.03	>.05
Rund-Sletto	28	5.17	7.55	26	1.85	6.81	7.64	<.01*
Bell-Social	29	3.28	5.95	26	0.00	7.43	8.77	<.01*

*Significant.

There is also some indication that the gain is greater for the members of the experimental group who were given the most guidance. The fact that the Social Preference scale shows an insignificant difference in mean gain for the two groups suggests that social guidance has an effect on the actual social behavior or amount of social activity but does not affect social preferences.

3. At the beginning and at the end of the experimental period the counselors rated the members of the experimental group on a rough scale of social adjustment. Only seven per cent of the group was rated lower on second rating and 38 per cent rated higher. There is no comparable measure for the control group, so the significance of this gain is difficult to interpret.

4. The experimental and control groups encircle about the same number of individual activities on retesting, but the experimental group encircles more group activities than the control group on retesting.

5. The experimental group reports more hours per week spent in extra-curricular activities than the control group and more offices and committees in these activities.

6. The experimental group indicates on a rating scale that they want to participate in fewer additional activities, think that they have made more friends, feel that they have participated in more activities compared with high school, and have a better opinion of the extra-curricular and social program on the campus than the control group.

7. Both the experimental and control groups feel that they are in fewer activities but have made more friends in college than in high school.

All of these findings combine to indicate that, from this small sample, social guidance and directed participation in extra-curricular activities improve the "social adjustment" of freshman girls as measured by personality scales and a questionnaire. Not only do the girls in the experimental group make greater mean gains, but they feel that they have more friends, participate in more activities, and are less critical of the social program than the control group. A treatment that makes people feel better satisfied with their social life is certainly worthy of further consideration. The problem was, however, essentially an investigation of a method and as such the results should be emphasized only as a justification for the further use of the method.

REFERENCES

1. Beaumont, H. "The Evaluation of Academic Counseling", *Journal of Higher Education,* X, (1939), 79-82, 116.
2. Brown, Clara. "A Social Activities Survey", *Journal of Higher Education,* VIII, (1937), 257-265.
3. Burks, F. W. "Some Factors Related to Social Success in College", *Journal of Social Psychology,* IX, (1938), 125-140.
4. Chapin, F. Stuart. *Extra-curricular Activities at the University of Minnesota.* Minneapolis: University of Minnesota Press, (1929).
5. Livingood, F. G. "Directed Extra-curricular Activities and Adjustments", *Mental Hygiene,* XX, (1936), 614-623.
6. Mallay, H. "A Study of Some of the Factors Underlying the Establishment of Successful Social Contacts at the College Student Level", *Journal of Social Psychology,* VII, (1936), 205-228.
7. Tuttle, H. S. "The Campus and Social Ideals", *Journal of Educational Research,* XXX, (1936), 177-182.
8. Williamson, E. G. and Bordin, E. S. "Evaluating Counseling by Means of a Control-group Experiment", *School and Society,* LII, (1940), 434-440.
9. Williamson, E. G. and Darley, J. G. "The Measurement of Social Attitudes of College Students. II. Validation of Two Attitude Tests", *Journal of Social Psychology,* VIII, (1937), 231-242.
10. Wrenn, C. G. *The Evaluation of Guidance,* Purdue University: Studies in Higher Education, No. 37, (1940), 51-61.

NEW TESTS*

Cooperative Chemistry Test for College Students, by B. Clifford Hendricks, B. H. Handorf, O. M. Smith, Chris P. Keim, Rufus D. Reed, Alexander Calandra, Ralph W. Tyler, and Fred P. Frutchey. Form 1942. Part I, Information and Vocabulary; Part II, Problems and Equations; and Part III, Scientific Method. Time, 90 minutes. 10 to 99 copies 6½c; 100 or more copies 6c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Cooperative English Test, by Geraldine Spaulding and Frederick B. Davis. 1942. Form S. Test A, Mechanics of Expression; Test B1, Effectiveness of Expression (Lower Level); Test B2, Effectiveness of Expression (Higher Level); Test C1, Reading Comprehension (Lower Level); Test C2, Reading Comprehension (Higher Level). Time, 40 minutes for each test. 10 to 99 copies 5½c; 100 or more copies 5c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Cooperative French Test, by Geraldine Spaulding, Laura Towne, and Sarah Wolfson Lorge. 1942. Form S, Lower Level for use in the first two years of high school or the first year of college; Higher Level for use with students who have had more than two years study of French in high school or more than one year in college. Part I, Comprehension; Part II, Grammar; and Part III, Civilization. Time, 80 minutes. 10 to 99 copies 6½c; 100 or more copies 6c; specimen set 25c. Part I available as separate booklet, 10 to 99 copies 5½c; 100 or more copies 5c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Cooperative Italian Test, by Peter Riccio and Anthony Cuffari. 1942. For students who have had two semesters or more of study of Italian. Experimental Form S. Time, 70 minutes. Part I, Reading; Part II, Vocabulary; and Part III, Grammar. 10 to 99 copies 6½c; 100 or more copies 6c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Cooperative Latin Test, by Harold V. King and Geraldine Spaulding. 1942. Form S, Lower Level to cover beginning Latin and Caesar;

---

*Prepared by Jane Gilbert.

Higher Level for use with students who have completed one semester or more of study beyond Caesar. Part I, Comprehension; Part II, Grammar; and Part III, Civilization. Time, 80 minutes. 10 to 99 copies 6½c; 100 or more copies 6c; specimen set 25c. Part I available as separate booklet, 10 to 99 copies 5½c; 100 or more copies 5c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Cooperative Test in Secondary School Mathematics (Higher Level), by Margaret Martin, William Mollenkopf, Radcliffe W. Bristol, William S. Litterick, and Carroll G. Ross. 1942. Form S. For grades 10 to 12. Time, 80 minutes. 10 to 99 copies 6½c; 100 or more copies 6c; specimen set 25c. Published by the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

Interest Inventory for Elementary Grades, by Mitchell Dreese and Elizabeth Mooney. 1941. Time, about 30 minutes. 5c each; manual 15c; specimen set 25c. Published by the Center for Psychological Service, George Washington University, Washington, D. C.

Meier Art Judgment Test, by Norman Charles Meier. Revised 1942. Grades 7 through adult. Time, about 45 minutes. Test books 75c; $3.50 for 5; $6.25 for 10c; 55c in lots of 25; record sheets 2½c; 2c per 100; manual 10c; sample set 90c. Published by the Bureau of Educational Research, State University of Iowa, Iowa City, Iowa.

Otis Classification Test, by Arthur S. Otis. Revised 1941. Forms R, S. and T. For grades 4 to 8. Time, 30 minutes for each part. Hand- and machine-scored. $1.25 per 25; specimen set 30c. Published by the World Book Company, Yonkers-on-Hudson, New York.

Pintner-Durost Elementary Test, by Rudolf Pintner and Walter N. Durost. For grades 2, 3, and 4. Form A, Scale 1 (Picture Content) and Scale 2 (Reading Content). $1.35 per 25 for Scale 1; $1.20 per 25 for Scale 2; specimen set (Scale 1 and Scale 2) 30c. Published by the World Book Company, Yonkers-on-Hudson, New York.

Preference Record, by G. Frederic Kuder. 1942. Form BB for self-scoring; Form BM for machine-scoring. For high-school and college students and adults. Time, about forty minutes. Test booklets 25c; answer pads 5c; profile sheets $1.25 per 100; specimen set 25c.

Published by Science Research Associates, 1700 Prairie Avenue, Chicago, Illinois.

Purdue Placement Test in English, by J. H. McKee, G. S. Wykoff, and H. H. Remmers. 1941. For high school seniors and college freshmen. Time, about 35 minutes. Form C, $1.65 per 25; separate answer sheets 75c per 25. Published by Houghton Mifflin Company, 2 Park Street, Boston, Massachusetts.

Terman-McNemar Test of Mental Ability, by Lewis M. Terman and Quinn McNemar. 1942. For grades 7 to 12. Time, 40 minutes. Forms C and D. $1.25 per 25; specimen set 20c. Published by the World Book Company, Yonkers-on-Hudson, New York.

Study-Habits Inventory, by C. Gilbert Wrenn. Revised 1941. For grade 12 and college. $1.25 per 25; $3.50 per 100; $2.50 per 100 for 1000 or more. Published by Stanford University Press, Stanford University, California.

Test of Practical Judgment, by Alfred J. Cardall. 1942. For 12th grade level and above. Time, about 45 minutes. Hand- or machine-scored. 10c each; specimen set 25c. Published by Science Research Associates, 1700 Prairie Avenue, Chicago, Illinois.

The World Test, by Charlotte Buehler and Gayle Kelley. 1941. To measure emotional problems. For clinical use with children 5 to 11. Time, about 20 minutes. Complete test materials, manual, and 25 record forms, $60.00. Published by the Psychological Corporation, 522 Fifth Avenue, New York City.

MEASUREMENT ABSTRACTS*

Bellows, R. M. "Procedures for Evaluating Vocational Criteria." *Journal of Applied Psychology,* XXV (1941), 499-513.

The fact that the basic vocational criteria used in the evaluation of predictive instruments are fallible is generally neglected. The source of fallibility may lie in such factors as (1) illicit use of predictive information giving previous knowledge of psychological test scores or other performance ratings; (2) artificial limitations of production brought about by physical conditions influencing output of work; (3) differential experience or training. To overcome the influence of criterion contamination several checks are recommended and evaluated. Knowledge of the future validity of a predictor is impossible because of various changes in the situation. No single procedure for criterion evaluation is adequate, which suggests that indices of validity are largely determined by the degree of fallibility of the criterion, and that the interpretation of such indices is dependent upon knowledge of the criterion used in validation. *L. Bouthilet.*

Buros, Oscar K. (editor) The Second Yearbook of Research and Statistical Methodology. Highland Park, New Jersey, Gryphon Press. 1941.

This yearbook has been compiled in an effort (a) to make students and teachers of statistics aware of inaccuracies and the inadequacy of much current statistical literature and information, (b) to serve as a source for selection of textbooks with discrimination, (c) to evaluate weak and strong points of statistical books, (d) to point out current developments in monograph and textbook writing and criticism, (e) to acquaint statistical workers with the broad applications of statistical work in many fields, (f) to present different points of view among students of statistical theory, (g) to improve the quality of such book reviews by more careful choice of reviewers and by stimulating reviewers not to review books which they cannot appraise adequately.

The editor has greatly increased the scope of this volume over an earlier one, including 1,652 review excerpts from 283 journals. An attempt has also been made to list books on research methodology in specific fields, although this list is by no means inclusive. However, this yearbook represents a significant contribution to the field of methodology and should make workers in this field more acutely aware of current developments. *Jane Gilbert.*

Cronbach, L. J. "An Experimental Comparison of the Multiple True-False and Multiple Multiple-Choice Tests." *Journal of Educational Psychology,* XXXII (1941), 533-543.

---

*Edited by Forrest A. Kingsbury.

Two subject-matter tests, one in multiple true-false form, and the other in multiple multiple-choice form were administered to 57 and 60 students, respectively. The former consists of multiple-choice items in which each alternative is marked true or false by the student; the latter, of similar items in which only correct alternatives are marked. Results showed the two forms were essentially equivalent. The hypothesis is advanced that the tendency to mark uncertain items "true" may be a personality trait which may influence the validity of true-false test scores. *L. Birdsall.*

Ewart, E., Seashore, S. E., and Tiffin, J. "A Factor Analysis of an Industrial Merit Rating Scale." *Journal of Applied Psychology,* XXV (1941), 481-486.

In order to determine how many traits actually influence the ratings, tetrachoric intercorrelations were computed for ratings on a twelve-trait scale constructed for use in a large industrial plant. This correlation matrix was factored by Thurstone's centroid method, and the factors rotated for simple structure. Three factors were obtained: I, a general factor, termed "ability to do the present job," accounts for most of the total variance of the scale. Factor II represents knowledge or skill over and above the requirements for the specific job. Factor III is on the variable "health." Factors I and III are orthogonal while Factors I and II are oblique. *K. S. Yum.*

Forlano, G. and Pintner, R. "Selection of Upper and Lower Groups for Item Validation." *Journal of Educational Psychology,* XXXII (1941), 544-549.

Two sets of data from the *Study Habits Inventory* and the *Home-Background Survey Test* have been subjected to item validation, using five different methods of selecting upper and lower groups. The authors conclude that for a simple and rapid, rough-and-ready method of validation of test items of the inventory type, the upper versus lower 27 per cent method is preferable, even though distributions are more or less non-normal. The other upper versus lower methods studied were 50 per cent, 33⅓ per cent, 16 per cent, and 7 per cent. *K. S. Yum.*

Gilman, W. A. and Gray, D. E. "Guessing on True-False Tests." *Educational Research Bulletin,* XXI (1942), 9-12.

The attempt to penalize guessing by subtracting the number of wrong answers from the number of correct answers is ineffective. This is clear from the study of a case in which there are n pure guesses. Theoretically the student would have $\frac{n}{2}$ correct answers and $\frac{n}{2}$ incorrect answers; hence the increment of $\frac{n}{2}-\frac{n}{2}$ would leave his grade on other

items unaltered. In practice, pure guessing rarely exists. The possession of partial knowledge gives the student better than a fifty per cent chance; hence it is to his advantage to guess on tests thus scored. *George W. Boguslavsky.*

Growdon, C. H. "The Revised Stanford-Binet Scale Applied as a Point-Scale." *Journal of Applied Psychology,* XXV (1941), 660-671.

Form L of the Revised Stanford (from Year VI up) has been arranged as a point-scale. By testing each subject only with that range of tests limited by 5 consecutive successes below the first failure and 5 consecutive failures above the last success, a very reliable mental age is obtained. Rescoring the records of 440 children of the usual clinical types yields I. Q.'s which correlate .976 with regular Stanford-Binet I. Q.'s. For eleven mental age levels the average saving in number of tests given was about 35 per cent, with I. Q. variations not to exceed 5 points in 9 of every 10 cases. *F. A. Kingsbury.*

Harding, J. "A Scale for Measuring Civilian Morale." *Journal of Psychology,* XII (1941), 101-110.

Out of a list of 59 items given to two criterion groups, high morale and low morale, 20 items were chosen to form the present morale scale. Each item is included in one of the four clusters: a. an attitude of confidence in the broad framework of capitalist democracy; its opposite, cynicism; b. an attitude of tolerance for various groups; c. an attitude of realism as opposed to wishful thinking; d. an attitude of assertive idealism in international affairs. Scoring for each item is on a five-point scale. Thus "total morale scores" may be computed. *Louise Grossnickle.*

Heston, J. C., and Cannell, C. F. "A Note on the Relation Between Age and Performance of Adult Subjects on Four Familiar Psychometric Tests." *Journal of Applied Psychology,* XXV (1941), 415-419.

Vocabulary Tests from Form L of the Revised Stanford-Binet Scale, Knox Cubes, Porteus Mazes and Ferguson Form Boards Tests were given to members of borrower families of the F.S.A. in Ohio, Maine, and Missouri. The data include 643 cases, 375 men and 268 women, all white. The age range for men was 15 to 76, and for women, 15 to 72, with medians at 37.5 and 35.0 years respectively. Two contrasting tendencies are noted on the age curves of scores of these tests. On the vocabulary test there is a rapid increase from age 15 to 20, then a slight rise up to 55, where a small drop occurs; while on the performance tests a rapid decline seems to be a characteristic tendency. *K. S. Yum.*

Jones, H. E. "Seasonal Variations in I. Q." *Journal of Experimental Education,* X (1941), 91-99.

A study of 19 comparisons of fall-to-spring versus spring-to-fall I. Q. changes in children of preschool age revealed that 18 of the 19 comparisons show a greater gain over the winter interval than over the summer interval. Four alternative hypotheses were considered:

1. Seasonal variations in the testers.
2. Seasonal variations in test performances.
3. The dependence of performance on seasonal variations in the child's activity.
4. The effect of seasonal variations on mental and physical growth.

George W. Boguslavsky.

Katz, Evelyn. "The Constancy of the Stanford-Binet I. Q. From Three to Five Years." *Journal of Psychology,* XII (1941), 159-182.

The Brush Foundation of Western Reserve has the records of 308 children of high socio-economic level, tested at six-month intervals from three to five years of age. "Test-retest correlations range from .533 to .765, the size of the correlations being unrelated to age but inversely related to the interval between tests." "The group as a whole shows a small increase in I. Q. with age." Large gains and losses of 20 or more points are more frequent over the longer intervals of time and for the younger ages. They are present in approximately 10 per cent of the test-retest comparisons, and occur for 40 per cent of the children. "These frequent fluctuations should probably be regarded as typical of children between three to five years who come from families of superior socio-economic status." *Helen M. Wolfle.*

Lindquist, E. F. A First Course in Statistics. Boston, Houghton Mifflin Company. 1941. 240pp.

This elementary statistics textbook presents a well-organized approach to the problem of measurement. An accompanying workbook has been designed to help the student integrate the theoretical approach with actual practice in applying these principles. The topics presented are as follows: frequency distribution, percentiles, graphical representation of frequency distributions, measures of central tendency, measures of variability, the nature of the normal curve, sampling error theory, standard measures and methods of combining test scores, correlation theory, and correlation techniques applied in the evaluation of test materials. *Jane Gilbert.*

Morrow, Robert S. "An Experimental Analysis of the Theory of Independent Abilities." *Journal of Educational Psychology,* XXXII (1941), 495-511.

"Eighty relatively homogeneous male college students were given

in a random manner" 23 subtests of standard tests of intelligence, artistic judgment, and clerical, mechanical, and manipulative ability. The correlations were analyzed by the "center of gravity" method into four factors. The factors were not rotated and were difficult to interpret. "By virtue of these findings," states the author, "it would appear that the Spearman and Thurstone theories are inadequate for explaining the relationships expressed in this study. Rather, one must conclude with the hypothesis that the abilities here tested are not disparate and static abilities, but that they are, instead, functional and dynamic relationships within the total personality." *Helen M. Wolfle.*

Reed, H. B. "The Place of the Bernreuter Personality, Stenquist Mechanical Aptitude, and Thurstone Vocational Interest Test in College Entrance Tests." *Journal of Applied Psychology,* XXV (1941), 528-534.

This investigation proposed to investigate the inter-relations between the Bernreuter, Stenquist, and Thurstone tests and scholastic achievement in order to find the place of such tests in a battery of college entrance examinations. The Bernreuter scores were also compared with teachers' ratings of traits of the same name as those in the test. Results show that there was little or no relationship between the three tests and scholastic achievement. It is concluded that the tests are of little value for guidance in choice of college courses, although the usefulness of the tests for other purposes was not investigated. *L. Bouthilet.*

Robinson, Frances P. Diagnostic and Remedial Techniques for Effective Study. New York, Harper and Brothers. 1941. 318pp.

This handbook has been evolved as a result of the author's experience at the State University of Iowa and an extensive how-to-study program at Ohio State University. The major emphasis in this book has been placed on diagnostic tests which are based on research analyses of college work and student errors rather than standard academic organization. The types of areas measured include study habits, reading skill, skill in use of academic resources, knowledge of fundamental processes and background knowledge, health, vocational planning, social adjustment, personal problems and motivation. All materials necessary for test administration and scoring are included in this book. Comparable retests are also available to help the student evaluate his improvement and to see the nature of his remaining problems. The book cannot be used independently by a student, but it should form a working basis for individual counseling and delineation of specific areas in which remedial treatment is indicated. *Jane Gilbert.*

Shuttleworth, F. K. "Sampling Errors Involved in Incomplete Returns to Mail Questionnaires." *Journal of Applied Psychology,* XXV (1941), 588-591.

There has been little attempt to determine the sampling errors due to incomplete returns of mail questionnaires. The only adequate check is to compare incomplete returns with complete returns. In a study of the employment status of certain university alumni, it was found that serious sampling errors were involved, the earliest returns coming from the more successful alumni. The conclusion is drawn that each questionnaire situation needs intensive study, which should include a complete return from at least a portion of the total population. *L. Bouthilet.*

Sloan, W. and Sharp, A. A. "A Note on Interpolation of Kent Oral Emergency Test Scores into Mental Age Years and Months." *Journal of Applied Psychology,* XXV (1941), 592-594.

The method consists of dividing equally the 12 mental age months in each year so that the first point at each year level falls exactly on that year. A corresponding column of I. Q.'s for adults is given with the chronological age of sixteen as a constant divisor. *K. S. Yum.*

Stalnaker, J. M. "A Note on the Computation of Y Values for Integral Values of X, when Y is a Linear Function of X." *Journal of Educational Psychology,* XXXII (1941), 559-560.

The author reports a method for rapid and accurate determination of converted scores for a large number of raw scores, with the aid of accounting machines and punched-card methods. The method demands a minimum of hand labor. The procedure is applicable to any situation where one set of scores is to be transmuted into any set of scores, providing the two sets are in a linear relationship, and the one variable changes in unit steps. *K. S. Yum.*

Super, D. E., and Roper, S. A. "An Objective Technique for Testing Vocational Interests." *Journal of Applied Psychology,* XXV (1941), 487-498.

A technique developed for testing vocational interests objectively is described. Pictures and films depicting different phases of various occupations are used. The assumption is made that memory for what is seen will be greatest in the field of greatest interest. Methods of validation are described. The test of interest in nursing was administered to 35 nurses and 111 high-school students, 36 of whom planned to enter nursing. Intelligence, previous knowledge, and success in nursing school influence the scores slightly or not at all. The present test showed no correlation with the *Strong Vocational Interest Blank.* The authors conclude the two are equally valid, but that the former measures degree of interest, whereas the latter compares the interests of subjects and those in the field. *L. Birdsall.*

Traxler, A. E. "The Reliability of the Bell Inventories and Their Correlation with Teacher Judgment." *Journal of Applied Psychology,* XXV (1941), 672-678.

Scores of 43 high-school pupils on the Bell *Adjustment Inventory* and the Bell *School Inventory* have been correlated by the split-half method. All the reliability coefficients are above .80, and some of them are close to or above .90. Correlations between the scores and the ratings by teachers and counselors on 33 pupils have been obtained. Four of the six correlations are statistically significant. However, the correlations, being low, fail to substantiate the validity of the inventories. The author suggests that we should have a criterion that will be much more defensible than a rating scale. *Louise Grossnickle.*

Yum, K. S. "Primary Mental Abilities and Scholastic Achievements in the Divisional Studies at the University of Chicago." *Journal of Applied Psychology,* XXV (1941), 712-720.

What particular combination of primary mental abilities is required for success in the divisional studies of the physical, biological and social sciences? The scores of 110 University of Chicago juniors were examined. "According to the critical ratios, there apparently exists no significant difference between the biological and social science groups." The mean profile of the physical science group (but not the total score) is significantly different from the other two. Induction distinguishes physical science men from biological science men, and deduction, space, and induction distinguish physical science men from social science men. The correlations of the factors with grades range from —.17 to +.52. "In general, the verbal, inductive reasoning, and deductive reasoning factors seem to correlate better with scholarship." *Helen M. Wolfle.*

MEASUREMENT NEWS

The Personnel Procedures Section, formerly the Personnel Research Section, of the War Department has developed in recent months a variety of classification, special aptitude, and achievement tests for the use of the Army. The section is currently interested in the selection of officer candidates and military specialists and in the training of physically and mentally limited men.

Among the officers on active duty with the section are Major Morton A. Seidenfeld, formerly of the National Tuberculosis Association; Lt. Donald E. Baier, on leave from the Mental Hygiene Bureau of the New Jersey State Hospital; and Lt. T. W. Harrell, who was in charge of research for the section in a civilian status. Captain Sidney Adams, formerly of the Employment Section of the Tennessee Valley Authority, has left the section for duty in the field. Among the civilian personnel of the section are: Dr. Clyde H. Coombs, formerly of the University of Chicago; Dr. Louise R. Witmer, on leave from Florida State College for Women; Dr. Bronson Price, formerly of Ohio State University; Dr. Reign H. Bittner, also formerly of Ohio State University; Mrs. Ruth D. Churchill, formerly of the University of Minnesota; Dr. Alvin C. Eurich, formerly of Stanford University; and Mr. Howard Uphoff, formerly of the U. S. Civil Service Commission.

Schools interested in building pupil morale for meeting war hardships will be interested in a "Test on the Effects of War" designed for the study of pupil morale and to identify war problems about which further instruction is needed. The test, prepared by Dr. Lee J. Cronbach, has been released by the School of Education of the State College of Washington, Pullman, Washington. The test is planned for grades 10, 11, and 12, but may be used at higher levels. Seventy statements about conceivable future developments are presented, and the pupil is required to respond by indicating how likely he thinks each effect is. Responses are analyzed to determine how optimistic or pessimistic each pupil is. Since good morale depends on a realistic outlook and planning for future developments, both the highly optimistic or complacent pupil, and the highly pessimistic, panicky pupil, are pointed out as cases for individual guidance. An item analysis of the responses of the group indicates those particular war problems about which pupils appear poorly informed.

The test is being made available as a professional service on a noncommercial basis to interested schools. For greatest value in planning the school program during wartime, the test should be given as early as possible. Question sheets, which may be used any number of times, sell for one cent apiece. Answer sheets, one of which is needed for

every pupil tested, sell for five cents each. This charge covers the cost of producing the test and of a complete scoring service. All papers are scored, analyzed, and interpreted by the State College without additional charge.

The test has been standardized on nearly two thousand pupils in the State of Washington, tested during January and February, 1942. The reliability of the Optimism score, on which the principal interpretations are based, is .77.

The Committee of Examinations and Tests, Division of Chemical Education, of the American Chemical Society, has announced that the 1942 Cooperative Chemistry Test will be available by April first. Inquiries should be addressed to the Cooperative Test Service, 15 Amsterdam Avenue, New York City.

The accumulation of data and experience in recent years has had the effect of modifying the concept of what the test should measure. As a result of extensive discussion at a conference held at the University of Chicago last June, the 1942 Form of the test is considerably different from the tests of the past four years. The test has been administered in a preliminary form to determine the difficulty and validity of each item. A brief description of the test follows:

Part I. *General Knowledge and Information.*

This section is based on knowledge of or *acquaintance with* important facts, definitions, laws and theories of chemistry. Historical events and application of chemistry to the social and economic world are represented.

Part II. *Application of Principles.*

This part attempts to measure the ability to solve numerical problems, to balance equations, and to make quantitative predictions by the application of chemical principles.

Part III. *Scientific Method.*

This section is concerned with the understanding of the relation of observation, definitions, laws, theories in the scientific procedure. The relation of theory to experiment is represented, as well as the ability to interpret chemical data.

Part IV. *Knowledge of Laboratory Technique and Procedure.*

This new section is included in the effort to measure acquaintance with the laboratory and knowledge of "correct" procedures. It does not attempt to measure skill or technique *per se.*

The committee which is sponsoring this test is comprised of the following members of the Division of Chemical Education:

B. Clifford Hendricks, University of Nebraska.
Rufus D. Read, New Jersey State Teachers College.
Ed. F. Degering, Purdue University.
Laurence S. Foster, Brown University.
Earl W. Phelan, Georgia State Womans College.
Theodore A. Ashford, University of Chicago.
Otto M. Smith, Oklahoma A and M College, *Chairman*.

The Annual Report of the Scottish Council for Research in Education states that a mass of records — 2,500 of scale L and 350 of scale M — have been collected with a view to standardizing the Terman-Merrill Revision of the Stanford Binet Scale for use in Scotland. It is hoped shortly to produce some evidence as to the suitability of this Revision for Scottish children.

A report on the follow-up of the random sample of 1,000 children and of the high scorers in certain counties who were given the Binet test in the 1932 Mental Survey is awaiting publication. It is interesting to note that an independent analysis of occupations has been made and correlated with each I. Q. group. The relation of occupation to age and to class on leaving school has been worked out with respect to both initial and final occupations; that is, to those occupations entered upon leaving school and to those held for not less than one year immediately before the close of the survey. A geographical analysis, based on the Four Cities, urban areas excluding the Four Cities, and rural areas, has also been made. The Annual Report states that so far as can at present be ascertained the correlation between intelligence and initial occupation does not appear to be very high, but this relation is closer by the time the occupation held at the close of the follow-up is entered.

Another report awaiting publication covers the results of an inquiry into methods of forecasting, at the qualifying stage, the pupil's later success. The methods considered are the traditional examination, scholastic tests, an intelligence test and teacher's estimate. It appears that the best combination, productive of the least number of misfits, is an intelligence test, an examination, and teacher's estimate "scaled."

The Psychological Classification and Research Sections of the Army Air Forces have established three Psychological Research Units. These units are located at Maxwell Field, Alabama, Kelly Field, Texas, and Santa Ana, California, and are headed respectively by Laurence F. Shaffer, Robert T. Rock, Jr., and J. P. Guilford.

All aviation cadets are given psycho-motor and group tests for the purpose of classifying them for various duties in the aircrew. In addi-

tion to administering these tests, the units do some research on the general problem of determining the aptitudes needed for different aircrew duties, as well as develop methods for the prediction of success in such duties.

These units are staffed by a group of officers and enlisted men. All the officers are well-qualified psychologists. Most of the enlisted men have done some graduate work in psychology and in addition have had some experience either in using psychological laboratory equipment or in the development, use, and validation of psychological tests. Periodically some qualified enlisted men are recommended for officer candidate schools. Successful completion of such schools leads to a commission.

Men interested in enlisting for such positions should send the following information to the Army Air Forces, Office of the Air Surgeon, War Department, Washington, D. C.: (1) full name, (2) date and place of birth, (3) local board number and order number, (4) four personal references, and (5) complete work and educational histories, including a detailed description of specialized training in psychology. Individuals who expect to be inducted into the service soon and who desire to be considered for assignment to work in psychology, should, in addition to the previous information, include (6) probable date of induction, stating whether notification of date of induction has been received and (7) probable place of induction.